

GenCore version 4.5
Copyright (c) 1993 - 2000 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:24:38 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196f-1
Perfect score: 25
Sequence: 1 gacattgcgcagtcacagaatla 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 199986

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : GenEmbl : *

1: gb_da : *

2: gb_hcg : *

3: gb_in : *

4: gb_om : *

5: gb_ov : *

6: gb_ph : *

7: gb_pl : *

8: gb_pr : *

9: gb_ro : *

10: gb_sy : *

11: gb_sy : *

12: gb_sy : *

13: gb_un : *

14: gb_vl : *

15: gb_da : *

16: em_hum : *

17: em_hum : *

18: em_in : *

19: em_om : *

20: em_ov : *

21: em_ov : *

22: em_pat : *

23: em_ph : *

24: em_pl : *

25: em_ro : *

26: em_sy : *

27: em_sy : *

28: em_un : *

29: em_vl : *

30: em_higo_hum : *

31: em_higo_inv : *

32: em_higo_rod : *

33: em_hig_hum : *

34: em_hig_inv : *

35: em_hig_rod : *

36: em_hig_other : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	6 AR029499	AR029499 Sequence
2	25	100.0	1041	6 AR034916	AR034916 Sequence
3	25	100.0	1084	6 A11530	A11530 B10B gene o
4	25	100.0	1121	6 E00893	E00893 Genomic DNA
5	25	100.0	5793	1 E00810	J04423 E.coli 7,8-
6	25	100.0	5872	6 A38246	A38246 Sequence 1
7	25	100.0	5872	6 A38251	A38251 Sequence 1
8	25	100.0	5872	6 A38251	A38251 Sequence 1
9	25	100.0	5872	6 A38251	A38251 Sequence 1
10	25	100.0	5872	6 A38251	A38251 Sequence 1
11	25	100.0	5872	6 A38251	A38251 Sequence 1
12	25	100.0	11022	1 AE000180	AE000180 Escherich
13	25	100.0	13501	1 AE005258	AE005258 Escherich
14	25	100.0	297816	1 AP002553	AP002553 Escherich
15	21	84.0	5526	1 AF250776	AF250776 Unculture
16	18	72.0	8227	1 AF248314	AF248314 Unculture
17	16	64.0	32767	2 AC087305	AC087305 Homo sapi
18	16	64.0	155172	2 AC021110	AC021110 Homo sapi
19	16	64.0	156184	2 AC041018	AC041018 Homo sapi
20	16	64.0	174815	2 AP001390	AP001390 Homo sapi
21	16	64.0	178056	2 AC091589	AC091589 Homo sapi
22	16	64.0	196550	2 AC020734	AC020734 Homo sapi
23	15	60.0	561	6 AX136747	AX136747 Sequence
24	15	60.0	1771	6 AR080743	AR080743 Sequence
25	15	60.0	1771	10 MUSINTESUB	L20788 Mus musculu
26	15	60.0	1829	11 G07181	G07181 human SYS W
27	15	60.0	1949	10 MINTESUB	U34627 Mus musculu
28	15	60.0	2340	1 ECHK31M	X82231 E.coli HK31
29	15	60.0	3140	6 Y029487S18	Y029487S18 Sequence
30	15	60.0	3520	6 AX136409	AX136409 Sequence
31	15	60.0	3748	6 AX098224	AX098224 Sequence
32	15	60.0	4602	9 D13644	D13644 Homo sapien
33	15	60.0	4636	1 AF039582	AF039582 Enterobac
34	15	60.0	5218	9 AF152315	AF152315 Homo sapi
35	15	60.0	5221	9 AF152315	AF152315 Homo sapi
36	15	60.0	5233	9 AF152315	AF152315 Homo sapi
37	15	60.0	5251	9 AF152312	AF152312 Homo sapi
38	15	60.0	5254	9 AF152306	AF152306 Homo sapi
39	15	60.0	5254	9 AF152310	AF152310 Homo sapi
40	15	60.0	5257	9 AF152307	AF152307 Homo sapi
41	15	60.0	5260	9 AF152305	AF152305 Homo sapi
42	15	60.0	5260	9 AF152309	AF152309 Homo sapi
43	15	60.0	5260	9 AF152311	AF152311 Homo sapi
44	15	60.0	5260	9 AF152314	AF152314 Homo sapi
45	15	60.0	5260	9 AF152316	AF152316 Homo sapi
46	15	60.0	5260	9 AF152317	AF152317 Homo sapi
47	15	60.0	5260	9 AF152303	AF152303 Homo sapi
48	15	60.0	5260	9 AF152304	AF152304 Homo sapi
49	15	60.0	5260	9 AF152304	AF152304 Homo sapi
50	15	60.0	5260	9 AF152304	AF152304 Homo sapi
51	15	60.0	5260	9 AF152304	AF152304 Homo sapi
52	15	60.0	5260	9 AF152304	AF152304 Homo sapi
53	15	60.0	5260	9 AF152304	AF152304 Homo sapi
54	15	60.0	5260	9 AF152304	AF152304 Homo sapi
55	15	60.0	5260	9 AF152304	AF152304 Homo sapi
56	15	60.0	5260	9 AF152304	AF152304 Homo sapi
57	15	60.0	5260	9 AF152304	AF152304 Homo sapi
58	15	60.0	5260	9 AF152304	AF152304 Homo sapi
59	15	60.0	5260	9 AF152304	AF152304 Homo sapi
60	15	60.0	5260	9 AF152304	AF152304 Homo sapi
61	15	60.0	5260	9 AF152304	AF152304 Homo sapi
62	15	60.0	5260	9 AF152304	AF152304 Homo sapi
63	15	60.0	5260	9 AF152304	AF152304 Homo sapi
64	15	60.0	5260	9 AF152304	AF152304 Homo sapi
65	15	60.0	5260	9 AF152304	AF152304 Homo sapi
66	15	60.0	5260	9 AF152304	AF152304 Homo sapi
67	15	60.0	5260	9 AF152304	AF152304 Homo sapi
68	15	60.0	5260	9 AF152304	AF152304 Homo sapi
69	15	60.0	5260	9 AF152304	AF152304 Homo sapi
70	15	60.0	5260	9 AF152304	AF152304 Homo sapi

c 71 15 60.0 122615 2 AP003965
c 72 15 60.0 123887 9 AL390739
73 15 60.0 128495 9 AL356913
74 15 60.0 131541 2 AC006594
75 15 60.0 137169 2 AC015781
76 15 60.0 138033 2 AP002339
77 15 60.0 143800 2 AC011875
c 78 15 60.0 143440 2 AC091542
c 79 15 60.0 145942 2 AC080001
80 15 60.0 146002 2 CNS01RGD
c 81 15 60.0 146847 9 AP000248
c 82 15 60.0 147172 9 AC008688
c 83 15 60.0 147706 8 AC083945
84 15 60.0 148189 9 AC018796
85 15 60.0 148582 9 AC015774
86 15 60.0 148845 9 HS1114A1
87 15 60.0 150613 2 AP003878
c 88 15 60.0 150893 2 AC044871
c 89 15 60.0 152719 2 AC012545
90 15 60.0 154244 2 AC073463
91 15 60.0 154547 2 AC016112
c 92 15 60.0 155627 2 AC013608
c 93 15 60.0 155822 2 CNS01DUV
c 94 15 60.0 156288 2 AC025647
c 95 15 60.0 158621 2 AC011789
c 96 15 60.0 159866 9 AC016598
c 97 15 60.0 160855 2 AC011579
c 98 15 60.0 161474 2 AC016198
c 99 15 60.0 162383 9 AC074045
c 100 15 60.0 164113 2 AL391820

ALIGNMENTS

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgaagtcacagaatta 25
|||||
Db 21 GACATTGTGCGAAGTCACAGAATTA 45

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgaagtcacagaatta 25
|||||
Db 21 GACATTGTGCGAAGTCACAGAATTA 45

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION Biot gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 1084)
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
source location/Qualifiers
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/gene="BIOB"
24..1064
/gene="BIOB"
/codon_start=1
/transl_table=1
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
/translation="MAHRPRMTLSQVTELEKPLDLEPAOOVHRQHPDPOYOVST
LSIKTAGCEDCKYCPOSSRYKLEAEELMVEQVLESARAKAAGSTRICMGAW
KNPERMMPYLEOMVOGVKMGLEACMTGLTISESAORLANGLDYVHNIDTSEF
VPTNMLKVGATPLADNDVDADFIPTIAVANIMPTSVYRISAGREQMOTQAMC
FMAGANSIEYGCILTTTPNEDEKDLQLEFRKILNPQYAVIAGDNEQQRLEQALMT
PDTDEYNNAL"

BASE COUNT 271 a 286 c 318 g 209 t

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgaagtcacagaatta 25
|||||
Db 44 GACATTGTGCGAAGTCACAGAATTA 68

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin synthetase.
PAT 29-SEP-1997


```

ACCESSION      E00893
VERSION        E00893.1 GI:2169154
KEYWORDS       JP 1986149091-A/1.
SOURCE         Escherichia coli.
ORGANISM       Escherichia coli
                Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
REFERENCE      1 (bases 1 to 1121)
AUTHORS        Hirono,Y., Kojima,T. and Kimura,H.
TITLE          DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
                PRODUCTION OF BIOTIN
JOURNAL        Patent: JP 1986149091-A 1 07-JUL-1986;
                NIPPON SODA CO LTD
COMMENT        OS Escherichia coli
                PN JP 1986149091-A/1
                PD 07-JUL-1986
                PF 24-DEC-1984 JP 1984272605
                PI HIRONO YOSHIIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
                C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18,PC
                C12R1:19);
                CC strandedness: Double;
                CC topology: Linear;
                CC hypothetical: No;
                CC anti-sense: No;
                CC *source: strain=Escherichia coli Ns101;
                CC Feature is identified by experimental;
                FH Key location/Qualifiers
                FT CDS 42..1079
                /product="biotin synthetase",
                /location/Qualifiers
                source 1..1121
                /organism="Escherichia coli"
                /db_xref="taxon:562"
                289 a 296 c 325 g 211 t

BASE COUNT    289 a 296 c 325 g 211 t

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1121;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacattgtcgcaagtcacagaatta 25
    |||||||
Db 62 GACATTGTGCGCAAGTCACAGATT 86

RESULT 5
LOCUS          EC0BIO 5793 bp DNA BCT 28-FEB-1994
DEFINITION    E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
                (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
                protein, and dehydrobiotin synthetase (bioD), complete cds.
ACCESSION      J04423
VERSION        J04423.1 GI:145422
KEYWORDS       7,8-diamino-pelargonic acid aminotransferase;
                7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
                bioC gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin
                synthetase.
SOURCE         Escherichia coli (strain K-12) DNA.
ORGANISM       Escherichia coli
                Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
REFERENCE      1 (bases 1 to 5793)
AUTHORS        Otsuka,A.O., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
                Johnson,O.
TITLE          The Escherichia coli biotin biosynthetic enzyme sequences
                predicted
JOURNAL        J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE        89066784
COMMENT        Draft entry and computer-readable sequence [1] kindly submitted by
                A.Otsuka, 09-NOV-1988.
FEATURES
                location/Qualifiers

```

source

CDS

gene
CDSgene
CDSgene
CDSgene
CDS

```

1..5793
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
/complete="98..574"
/feature="orf 1"
/codon_start=1
/transl_table=1
/protein_id="AA23513.1"
/db_xref="GI:455168"
/translation="MKLISMDLFDGKRLHRRHVGNGMGYDGNISPHLAMDVPAGTK
SFVVTGYCDPAPPGSGMHWVNVNLPADRVLDYDGGSGSLVAMPDGYLOTTRDEKGTG
YDCAAPPGGTHRYFVHALDIERIDYDGAASAAWGVNHFHSLASITAMS"
/complete(633..1925)
/gene="bioA"
/complete(633..1925)
/gene="bioA"
/complete(633..1925)
/gene="bioA"
/codon_start=1
/transl_table=1
/transl_table=1
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AA23514.1"
/db_xref="GI:457106"
/translation="MTDDLAFDGRHIMHPTSMTPSPYPVYSAEGCELISDGR
LVDGMSWMAITHCYNHPQLNAAMKSOIDAMSHVMEGITHAPALIELRLVMSGRN
ALECVFLADSGSVAEVAMKALQYQAKGEARQRLTFRNGYHGDFPGMSYCDPDN
SMHSLMKGYLPEENLEFAPAPQSRMDGEMWDERDMGKPARLMAHHEHETAAVILIEYDGA
GMRMYHPEMLKRIKICDREGILLIADLETATGFGKLPACHEHETIADILICLKA
LTGGMTLSATLTTRREVAETISNGEACFMHGPFPKGNPLACAAANSLALIESGMQ
QOVDIEVQLREQLAPARDAEMVADVRLGAIQVETTHVNNAAALQKFEVEGVIR
PFGKILYLMPTIILPQOLQRLTAANNVAQDDEFQCO"
2012..3052
/gene="bioB"
/complete(633..1925)
/gene="bioB"
/complete(633..1925)
/gene="bioB"
/codon_start=1
/transl_table=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AA23515.1"
/db_xref="GI:145425"
/translation="MAHPRWTLISOVTELEFEKPLDLLFEAOVYHROHDPROYVST
LSIKTGACPECKKYCPQPSRYKRTGIEERLEMEVOVLESARKKAAGSTFCGAAW
KNHEDRMPLYLEDMOVGVKAMGLEACMTLGTISEQAOARLANAGIDLYNHLIDSPER
VGNILITTRYOERLDTLEKVRDAGIKVSGIVGIGETVVKRAGLLQLANLPTPES
VPTNMLVKYKGPLADNDVDADFRTIAVARIMMPTSYVLSLGRQNEOYQALC
FMAGANSTYCGKLLTTPNPEEDKDLQLEFRKGLMPQQTAVLADGNEQGRLEQALMT
PDTDEYNNAAAL"
3049..4203
/gene="bioF"
/complete(633..1925)
/gene="bioF"
/complete(633..1925)
/gene="bioF"
/codon_start=1
/transl_table=1
/transl_table=1
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AA23516.1"
/db_xref="GI:145426"
/translation="MSWQEKINALLDARRAALRRRYVAVOAGACRMVADPROYVNF
SSNDYLGISHHPOIRAMQGAEPFGIGSGSGSHSGYGVVHQALEELAEVLGSKRA
LLFTSGAANQAVIAAMAKEDRTIADRLSHSLLEAASLSQSLRRFAHNDVTLAR
LLASPCGQOQMVYTEGVGSMGDSAPLAETIQOVTQOHNGLMWDAHNGTGVGEOGRG
SCWLOKVRPELVYTRGKGVGCAAVTCSTYADVLLQFARHLIYSTMPPOAOAL
RASLAVIRSDGDAERREKLAALITRRPAGQDLPPTLADSCSAIQPLIVGDSRALQL
AEKLRQOGCWTAIRPTVPAGTARLRLTLTAHEMODDRLLEVLHGNG"
4190..4945
/gene="bioC"
/complete(633..1925)
/gene="bioC"
/complete(633..1925)
/gene="bioC"
/codon_start=1
/transl_table=1
/transl_table=1
/db_xref="GI:145427"
/translation="MATVNRQATIAAFGRAAHYEHQADLQROSADALLAMPQRYT
HYLDAGCGGMSRHRHREHQAQVALDSPMIVQAKQDAADHYLAGDIESLPLATA
TFDLMSNLAVQMCNISTALRELRYVRVRKGVVAFPTLVQGSILPERHQAQVADBRP

```

```

HNRFLPDEIEOSLNGVYOHIIQPIITLWEDLASMRSLKIGATHLHGDRPKIL
TTSQLORLQILAMPQOQGRYPLTYHLFLGVIARE"
gene
4938..5597
/gene="B10D"
4938..5597
/gene="B10D"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYEVTGTFEFGKTVASCALLQAAKRGATYKPVASGS
EKTEGLRNSDALQORNSLQDYATVNVYTFAPSPHILISAGCRPIESLVSAG
LRALEQOQDWLVLEGAGWFLPSLDPTTFADWDQOLPYLVGVKLCQINAMLIYA
QVITAGLITLAVGANDVTPGKRHAETMTTLRMIPRCWEKSPGLQIKMROPES
T"
BASE COUNT      1363 a 1554 c 1631 g 1245 t
ORIGIN          4626 bp upstream of HpaI site: 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 1; Length 5793;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gacattgcgcgaatcacagaatca 25
Db      2032 GACATTGTCGCAAGTCACAGATTATTA 2056

RESULT 6
LOCUS      A38246      5872 bp      DNA
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION  A38246
VERSION     A38246.1 GI:2294844
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
REFERENCE   1 (bases 1 to 5872)
AUTHORS     Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE       BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     LONZA AG (CH)
COMMENT     other publication PL 308301 950724
            other publication CA 2145400 940414
            other publication AU 4820293 940426
            other publication HU 71781 960228
            other publication SK 42095 951108
            other publication CZ 9500809 950913
            other publication FI 951547 950331
            other publication JP 8501694T 960227.
            location/Qualifiers
FEATURES
    source
        1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PB030A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..28
            /evidence=experimental
            45..49
            /standard_name="PROMOTER PTAC"
            /evidence=experimental
            105..119
            /standard_name="B10B RBS NO.9"
            /evidence=experimental
            117..1157
            /gene="B10B"
            117..1157
            /gene="B10B"
            /gene="B10B"

gene
117..1157
/gene="B10B"

CDS
/gene="B10B"

/codon_start=1
/transl_table=11
/product="B10C RBS"
/protein_id="CAA02326.1"
/db_xref="GI:2294845"
/translation="MAHRPRWTLQVTELEFRLDILFEAQGVHROHFDPROGVST
LSTKTACPEDCKYCPOSSRYKTGLEARIEMVEBOVESARAKAAGSTRGCAAR
KNRBERMPYLEQMOVGVKAMGLACMTGTLSESOARLADYNNHNDTSPEP
VGNITRTYQOEHLDTLEKVRDAGIKVCGGIVGIGETKPKDAGLIDLANIPDES
VPINMLVKGVTGLADNDVDADPDLRTTAVARIRMPSTYVLSGRQDMQEQVAMC
PMDEXYENMALT"
2284..2297
/standard_name="B10C RBS"
2295..3050
/gene="B10C"
2295..3050
/gene="B10C"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATYVKQAIATAAFGRAAHYBOHADLQROSADALLAMPQRYT
HYLDACCGGKMSRHRERHQAOTLALDLSPPMLVQARQKDAADHYLAGDIESLPATA
TFDLAWSNLAVQWCGNLSTALRELRYRVRPQGVVAFPTLVQSLPELHOAQVDERP
TTSQLORLQILAMPQOQGRYPLTYHLFLGVIARE"
3742..3752
/standard_name="B10A RBS"
3750..5039
/gene="B10A"
3750..5039
/gene="B10A"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDLAFLPORHHTMTSPSLPYVYVVSAGECELLSDGR
LVDSSMMMAATGCHYNNHPOLNAAMKSIDAMSHVMFGITPAPELCKRLVAMPPOP
LECFELDSSVAEVAEMKRALIQWQAKGEMWDERMVGFAALMAHHEHIAAVIIEPYGAG
GMRYTHPEWLKRIKICDREGILLADELATGFGKPLFACEHIAIDILCGKAL
TGGTMTLSATLITREVAETISNGEACFPMTGTPMGNPPLACAAASLALLESQMOO
QVADIEVOLREQLAPARDAEMVADVAICIVTCTTHVYNNALQKFFVEQGWIRP
EGKILYMPPIYILLPOQLQRLTAANRAVODETFPCQ"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6
/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRQGDKLPHRHVFNKGKYGCDNITSPHLANDVDVPAATK
SFVVCYDPAATGSGMMHWVYVNPADTRVLPGCFSSGLVAMPDGLQTRTDGKKG
TGAAPPKGETHRYITFVHALDTERIDVDGASGAMGCFVNHFSIASASITAMFS"

```

stem_loop 5583..5605
terminator 5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacatttcgcaagtcacagaattta 25
|||||
Db 137 GACATTTCGCAAGTCACAGAATTTA 161

RESULT 7
A38251 5872 bp DNA PAT 05-MAR-1997
LOCUS Sequence 6 from Patent WO9408023.
DEFINITION A38251
ACCESSION A38251.1 GI:2294849
VERSION A38251.1
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
location/Qualifiers

FEATURES
source 1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PRO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=1
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQKRTINMALDRAADLRARRRYPVAGCAGRLVADDDROYLNF
SSNDYLGISHPQITRAMOGADDFGSGSGHVSQSVYHQALEELAEMLGYSSRA
LLEISGEFANQAVIAAMAKEDRTAARLSHASLLEASLSPSOLRRPAPHNVTHLAR
LLASPCDQGMVVEGVSMGDSAPLAEIQVYTOOHNGMLVDDAHGTCVGEORC
SCMLQKVPDLVTFGKFGVSGAAVLCSTVADYLLQFARHLITVSTMPROQAL
RSALAVIRSDGDAARREKLALLITFRAGVODLPFTLADCSAIOPLVDNSRALQTL
AEKIQGSCWTAIRPPTVPAGTARLRLLTLAAHEMODIDRLLEVLHGNC"

RBS 3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"

CDS 3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=1
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRPVTCGTDFEVGKYVASCALLQAAKAGCTRTAGTYPVAGS
EKTEPGLRNSDALAQNSSIQLDYAVNPVYFAEPTSPHIIISADGRTESLVNSAG
LRAIEQADWVIVGAGCWFTPLSDTFTEADWYTOQLPVILVGVKGLGCIHHALTA
QVIOHAGITLTAGVANDVPPGKRHAHEYMTTLRMIPAPLLGEIPIWLENPENATYK
YINLAFVDASTLGTSTRL"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacatttcgcaagtcacagaattta 25
|||||
Db 137 GACATTTCGCAAGTCACAGAATTTA 161

RESULT 8
A93674 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 1 from Patent EP0798384.
DEFINITION A93674
ACCESSION A93674.1 GI:6741862
VERSION A93674.1
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
location/Qualifiers

FEATURES
source 1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PRO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=1
/evidence=experimental
/product="BIOFIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPRTVLSQVTELFERKPLDLLPEAQQVHROHPPROYVST
LLSIKTGACPEDCKYCPQSSSKYKTGLAEIRLMEVYLESARAKAAGSTRCKMAAW

KPHREMPYEQWGVKAMLEACMTLGLTSESQORLANAGLDYYNHNLDTPSPF
IGNLITRTYQERDITLEKVDAGIKVCSGIGTGTGVDKDRGLLQLANLPTPES
VPIIMLVKVGKPLADNDVDPAFDEIRTAIARIMPTSVRLSAGEQNRPTQAMC
PMAGANSIFYGCKLTTPNPEKDLQLFKRLGLNPGQTAVLAGNDNDOORLEOALMT
PDTDEVNAAL
2284..2297
/standard_name="BIOC RBS"
/gene="BIOC"
/gene="BIOC"
2295..3050
/gene="BIOC"
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMLOX-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVNRQALIAAFGRRAAHYEQHADLQROADALLAMPQRYT
HYLDAGCGPMGRSHRERHAQYALDISPMIYQKQKADADHYLAGTESPLATA
TEFLAMGNLAVOMGNSTALRELKYVPRKGYATTLLVQSLPELHQWQVDEKP
HANFELPDELESLNGVHQHHIQPTITMLFDALSMKSLGIGATHLHEGRDEKL
TRSGQLRLQALAMPQOQGRPLTYHLEFLGYARE"
3742..3752
/standard_name="BIOA RBS"
/gene="BIOA"
/gene="BIOA"
3750..5039
/gene="BIOA"
/gene="BIOA"
3750..5039
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANONATE AMINOTRANSF."
/BC_number="2.6.1.62"
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDRHIMHPYTSMTSPLYVYVVAEGCELLSDGR
LVNGSSMMALIGHVNHQINAKMSQIDAMSHVFCGITHAPALECRKLVAMTPOI
LBCVPLADSGSAVEAVAMKALQYQAGKGAORFLTRNGYHGDTPGAMVCDPDS
MSLMKGYLPENLFAPOSRMDCDERMDYGFARLMAARHETIAVLIPELVQAG
GMRMYPEWLKIRKICDREGILLADELAGSPGTGCTACETAETAPDLICGKAL
TGGTMTLSATLTREVAERTISNGAGCGMGTPTGNPLACAAASLALIESGDWQ
OYADIEVOLREQLAPARAEYADRYGATGVETTHPVMAALQKFEQGWAIIP
FKLILMPYITLPOQLORLTAANRAVODETECO"
5088..5100
/standard_name="ORF1 RBS"
/gene="ORF1"
/gene="ORF1"
5098..5574
/gene="ORF1"
/gene="ORF1"
5098..5574
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNLDKRGDULPRHVFNGMGYDGDNI SPLLAMDVPACTK
SPVVTGYPDPATPGSGMMHWVNLPADEVLPDGFSGSLVAMPDGLQTRIDFGKTG
YDGAIPKGETHRYITFVHALDIRIDVDEGASGAMGVNHFHSLASSTAMPs"
5583..5605
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 137 GACATTGTCGCAAGTCACAGAATTA 161
RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
Source
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PRO30A15-9"
1141..1156
/standard_name="BIOF RBS"
/gene="BIOF"
1154..2308
/gene="BIOF"
1154..2308
/standard_name="8-AMINO-7-OXONONANONATE SYNTHASE"
/BC_number="2.3.1.47"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAAIDARADADLRRTPYAGAGRMVLADROYNT
SSNDYGLSHHPQITRAMQCAEQFGSGSGSVSYVHQALEELAEWIGYSRA
LLFTSGFANQAVITAAAMAKEDRIAORLSHASILEASLSPQLRRFANVTILAR
LLASPCGQQWVTEGVSMGDSAPLAEIQVYQINGMLAVDDAHGTVIGEORG
SCHLQKVPDELLVTFKRGFVSGAVALCSYADYLLQFARHLIYSTMPPAQOAL
RASLAVIRSDGDAAREKLAALITRFAGVQDLPFTLADSCSAIOPLIYDMSRALQL
AEKLRQGCWTAIRPPTVPAGTARLRLITLTAHEQDIDRLLEVLHGNG"
3030..3045
/standard_name="BIOB RBS"
/gene="BIOB"
/gene="BIOB"
3043..3753
/gene="BIOB"
/gene="BIOB"
3043..3753
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYPVGTDEVGKTVASCALLQAAKAAGYRRAGKRPVAGS
EKPEGLRNSDALQONSLQIDYATVTPYPAETSPHITSAOGRPISLVNSAG
OYDORACITLITAGVANDVTPGKRRAEYMTLITRMIPAVLLGEIPWLAENBNMATGK
YINDAEVDASTLGEFTRL"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaatta 25
 ||||||||||||||||||
 Db 137 GACATTGTCCGACAGTCACAGAATTA 161

RESULT 10
 ARI01809 5872 bp DNA PAT 14-FEB-2001
 LOCUS ARI01809
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES
 source 1..5872
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaatta 25
 ||||||||||||||||||
 Db 137 GACATTGTCCGACAGTCACAGAATTA 161

RESULT 11
 ARI01810 5872 bp DNA PAT 14-FEB-2001
 LOCUS ARI01810
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES
 source 1..5872
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaatta 25
 ||||||||||||||||||
 Db 137 GACATTGTCCGACAGTCACAGAATTA 161

RESULT 12
 AE000180 11022 bp DNA BCF 01-DEC-2000
 LOCUS AE000180
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete
 genome.
 ACCESSION AE000180 U00096

VERSION AE000180.1 GI:1786988
 KEYWORDS Escherichia coli K12.
 SOURCE Escherichia coli K12
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett, G., III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software. Kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@amber.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). ***The E. coli K12 sequence and its annotations are periodically updated: this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES
 source location/Qualifiers
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 <1..130
 /note="REP (repetitive extragenic palindromic) element; contains 4 REP sequences" complement(147..1430)
 /gene="ybhC"
 /note="b0772"

repeat_region
 gene

CDS	complement(147..1430) /gene="ybhC" /function="orf; Not classified" /note="f429; 98 pct identical to fragment YBH_ECOLI SW:P46130 (300 aa) but contains 127 additional C-terminal residues" /codon_start=1 /transl_table=11 /product="putative pectinesterase" /protein_id="AAC73859.1" /db_xref="GI:1786989"
promoter	/translation="MTTFSVRLALAFVGLTACSSNPDPDQPSDONAPQTSSKRPISAKAQNEDKHQYFASLTGCAAMNSPTILLPAODPEVVGAGTGCTGTHHTTIOAAVTAAAIIRTKRKYDAIMPEGEYGVYPAARGGGITITGTCEKPIDXKTGISLDGGSGSTDMRWYDRNGRYKMPGKPKAWMYDYSCOSCRSDSICVCASAFMSQNNGLQLOLWLTENNYIGDDVDIVSGCGAVFYDTBTRVVNSRTIQEAAVEAPATLSNLYGFTLVNRFNFANGDEVNOLGSLDVNDANTNQVTRISAINEGFNTAKPMADAVISNRPFAGTGSVLDDDELIDNTJNRRMKEVNNRGYSKVVAEAK"
gene	/note="factor Sigma70" complement(1507..1535) /note="Factor Sigma70; predicted +1 start at 806574" complement(1592..2058) /gene="ydhB" /note="b0773"
CDS	complement(1582..2058) /gene="ydhB" /function="orf; Unknown" /note="f158; 99 pct identical to ydBH_ECOLI SW: P12994" /codon_start=1 /transl_table=1 /product="orf, hypothetical protein" /protein_id="AAC73860.1" /db_xref="GI:1786990"
proteio_bind	/translation="MKRLISDLRDGDKLPHRHVEFGMGVDCDNISPLHAMDDVPAGIKSFVTCIDPDPAIPGSGMHMVVNLPADTRLVLCQFCGSLVAMPDGVLIOTRDEGRKGYDAAPRKGGTHRIETVALDIERIDVDEGASGMVFNHFILASIASITVMFS" /note="central position to predicted promoter:85.5" /bound_moiety="phas predicted site" complement(2117..3406) /gene="bioA"
gene	/note="b0774" complement(2117..3406) /gene="bioA" /RC_number="2.6.1.62" /function="enzyme; Biotynthesis of cofactors, carriers: Biotin" /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995" /codon_start=1 /transl_table=1 /product="7,8-diaminopelargonic acid synthetase" /protein_id="AAC73861.1" /db_xref="GI:1786991"
CDS	/translation="MTTDDIAFDQRHIWMHYTSMTSPLPVYPVNASGCCLLISDGRLVDGSMMAALHGHTNPQLNAARKSQIIDASHYMHCGLIHARAILCKRLVAMPPOPLECYVLASGSAVEAVAKKMALQYKAQKGARORFLIFRNQYGDFFGMASVCDPPNSHMUKWGHLPRNLFAFPARQMDDEMDEPDMVEGARLMAAHREHIAATVTPYOGAGTGGTWLRSATLTTEVEATTISNGAGCGFMHGPFRKNPLACAAANSLATIESGMQQOVARLEVOLRQDALPARDAEMADVARYLKATIGVETTHPVNMAALOKFEVEGWVIPFGKIITYMPPIITLPQOLRIRTAAYNAVODEVFQCQ"
promoter	complement(2193..2221) /gene="bioA"
proteia_bind	/note="factor Sigma70; predicted +1 start at 807260" 3411..3450 /note="central position to bioB promoter: -20"
proteio_bind	/bound_moiety="BioB documented site" 3411..3450 /note="central position to predicted promoter: -20; genetic evidence for the site" /bound_moiety="BioB predicted site" complement(3411..3450) /note="central position to predicted promoter:50"
proteio_bind	

protein_bind
 /bound_moiety="B10B predicted site"
 /complement(3411..3450)
 /note="central position to b10a promoter:50"
 /bound_moiety="B10B documented site"
 3413..3441
 /note="Factor Sigma70; promoter b10b; documented +1 at
 808525"
 complement(3447..3473)
 /note="Factor Sigma70; promoter b10a; documented +1 at
 808515"
 3493..4533
 /gene="bioB"
 /note="b0775"
 3493..4533
 /gene="bioB"
 /EC_number="2.8.1.1"
 /function="enzyme; Biosynthesis of cofactors, carriers;
 Biotin"
 /note="o346; 99 pct identical to B10B_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRRPMTLSQYTELEFEKPLDLLFEAGQVROHFPDPROYST
 LLSIKTGACPEDCKYDPOSSRYKRTGLEAPRLAEVROYEESARKAAASTPFGCAAMA
 KNERHMDPEYLEOMVGVKAMGLEACMTLGTLSFSDARLNACGLDYNHLIDTSPPEE
 VGLITRTYOEHLDDTLEKVRDAGIKVCSGGYIGETV KDRAGLLQLANLPPPEE
 YPIFNMLVVKYKGTPIADNDVDAPDFIRTIAYARIMPTSYRLSGRQNEOYQAMC
 PMGANSIFYGCKLITTPREDKDLQLPKILGIMPQOTAVLADNBOQRLFOALMT
 PDDEYYNAAL"
 4530..5684
 /gene="bioF"
 /note="bioF"
 /EC_number="b0776"
 4530..5684
 /gene="bioF"
 /EC_number="2.3.1.47"
 /function="enzyme; Biosynthesis of cofactors, carriers;
 Biotin"
 /note="o384; 100 pct identical to B10F_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSNDEKINAAIDARRADALRRRYPAVAGAGRWLVADDROYLNF
 SNNYDGLSHRPQIIRAKOQCAEQFGIGSGSGHVSQSYVHQALEBEELAEWIGYSRA
 LLFISGFANQAVIAAMAKEDRIADRLSHASLLEASLSPQLRRFANDVTGLAR
 LLASPCGQMWVTEGVFSMDGSAPLAEIQQYTOOHNGLAVDDAGHGVIGGEORG
 SCSMLQKAPPELLVTFPGKFGVSGAAVLCSSTVADYDLDLPARLLIYSTMPRAQNAL

Query Match 100.0%; Score 25; DB 1; Length 11022;
 Best Local Similarity 100.0%; Pred. NO. 1.3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gacatctgcgcgaagtccagcaaat 25
 ||||||||||||||||||||||||||||
 Db 3513 GACATTGTCGCAAGTCACGAGATTA 3537

RESULT 13
 LOCUS AE005258
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 of 155.
 ACCESSION AE005258 AE0051174
 VERSION AE005258.1 GI:12513751
 KEYWORDS
 SOURCE Escherichia coli O157:H7 EDL933.
 ORGANISM Escherichia coli O157:H7 EDL933
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 13501)

AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., Dimantola, E., Potamou, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.

TITLE Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7

JOURNAL Nature 409 (6819), 529-533 (2001)

MEDLINE 21074935

REFERENCE 2 (bases 1 to 13501)

AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., Dimantola, E., Potamou, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source location/Qualifiers

1..13501

/organism="Escherichia coli O157:H7 EDL933"

/strain="EDL933"

/serotype="O157:H7"

/db_xref="taxon:155864"

/note="enterohaemorrhagic"

11..7576

/note="O-island #36; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933K; includes one copy of the 13 bp direct repeat that flanks the prophage"

66..665

/gene="lomK"

/note="20981"

66..665

/gene="lomK"

/function="putative membrane: Other or unknown (Phage or Prophage Related)"

/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from Genpept 118 : g115732789|gb|AA653231.1|AF151091_2 (AF151091) lom [prophage P-Elba]"

/codon_start=1

/transl_table=11

/product="putative outer membrane protein of prophage CP-933K"

/protein_id="AAG5137.1"

/db_xref="GI:12513752"

/translation="MRKVCATILSAITCAVSGVPAMASEHOSHTLSAGYLAHSTDPAGSDLLINIKVRIEPTDTGLITSTSYANAEDKOTHSDDPMHEDYVRNHPFSVMAGPSYRVNWFSAVAMAGVATSVSTSGDIFRVTDNKRKTHDVLITGSDARTSNSTSLAN GAGVQNPTEVSADVAVEXSGSGDMRTDGFIVGVGKRF"

719..2041

/gene="20982"

719..2041

/gene="20982"

/function="putative structure: Structural component (Phage or Prophage Related)"

/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from Genpept 118 : g114585436|gb|AD25464.1|AF125520_59 (AF125520) putative tail fiber protein [Bacteriophage 933w]"

/codon_start=1

/transl_table=11

/product="putative tail component of prophage CP-933K"

/protein_id="AAG5138.1"

/db_xref="GI:12513753"

/translation="MGMAAVOISGLTKDCAKPIQNCITQIKARNSTVAVVTVASENPDAGRSMDVEYGVSVTLVIGGPPSHAGTISYEDSOPGTLDNDFLGAMTEDARPEALRREPQWYEEAARHAEEAKKAGCAETASAGISAKSAANADVTSAEDASE SARQAEASASAKRS"ASSSSASEAQAQKRSSELSQSTDAEISKRTAESACGMAADA TTSTKARESAESAQSAEQSRIAEADNRIPTVVGPPKGGEPGAGQGRGDKGE

gene 2489..3469

CDS /gene="20985"

2489..3469

/gene="20985"

/function="orf; Other or unknown (Phage or Prophage Related)"

/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from Genpept 118 : g11690367|gb|AA93527.1| (AF170176) hypothetical protein predicted by Glimmer [Salmonella typhimurium LT2]"

/codon_start=1

/transl_table=11

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AAG5140.1"

/db_xref="GI:12513755"

/translation="MLSPRTTFHNSVNIVQSSPCQVSPAGKEKELKVIDEPTLFEQMFENPERYKDEVPVNTKQHPYLDVNTAARIESDRMIGTIVDDESVNKRKAFS KLEDFENMYILYREDVDFSMYDRKLDSDIYADIDICEORLRTGDRDYLNLAKELR FISKADSLISMYAKKRHHAWDFERNALAKGEIFRCVYNTKNGHISGEGCTIYD MMMLTGKIGTIVAPDGISMHYDRNDSVNTENSAIIVNRSNHPALLEGISFMHSKYD AHPYDGLCKGYKKKFNFTPLNHNHFCDFIEPHNPNTIMNTSQYTCSSW"

3530..4522

/gene="20986"

3530..4522

/gene="20986"

/function="orf; Other or unknown (Phage or Prophage Related)"

/note="No significant matches"

/codon_start=1

/transl_table=11

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AAG5141.1"

/db_xref="GI:12513756"

/translation="MKIPSLQSNFNSAPAGYSAPDIAPNAENAVDYLDIGRRIPL SADDLVNYESVTRAYHDSRPLDIOHIVDMIGNVLALDSOTFRFVAVSYGLINEK VHTGCTKRYNEYELNEESSVKTDDIOSLTLKXVYDGOEPIFPCAGGNDDEPY VSFSAVPTDTSYEMPMSWOGGLIHETIIHHVTGSSDSDSGSNTILGETPILARVQOEFG WSYVDFKGYAEPEEREAHLRLRNALRQAAMHEENRFAFERLSTIDRDEASDFDT EYSVNSIGYGFIOQHDFFGLAINDMLADANOIOLYHGAPLYTFEGDVKHNOQ"

5350..6231

/gene="20989"

5350..6231

/gene="20989"

/function="orf; Other or unknown (Phage or Prophage Related)"

/note="No significant matches"

/codon_start=1

/transl_table=11

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AAG5142.1"

/db_xref="GI:12513757"

/translation="MLSPYSVNLGCSWNSLTRNLITSPDNKRVLSVYRDAAVHSDNGAQV

```

KVGNRTRYRVAVTDNKNFCVYTPRESHGCFNTLHLRLGMPKGRISRTKLEVNINAPSVAAK
BEGIVASNRNDLPDYADPPLPSVDYDNRISVPCNNVCKGKNNAVVEDADATKVIK
PTSGSNEEVTSEVGFENQYXGASSAKETIGNNDITIGIMDKTINGSLNLSLPAQ
AEHAIYMDPDLBOKGILFVDTTETNVLDPARKEFNPIDISSVNWSDRSMSQIMQ
SYHGSKDPLISVLSKI"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG55143.1"
/db_xref="GI:12513758"
/translacion="MRPSTLNLVHOSRSSSDTDIESLVKASSVQWIKNNPOLRF
OGTHNITQOIEALDKIGSTENGRLVLANIESIRLKESEVIVIHINSRLGVMARH
IDAEHNGTSDPHCNINAVPCGEGISVDPHATIFVHELHVPNNGERLKE
SRASOKYSPDLLEAFATVGLAFSEEVLENKFEHEICMPRTSPXSALIHDDNT
VSLGFQOVRHLPL"
complement(7668..8144)
/gene="ybhb"
/note="20992"
complement(7668..8144)
/gene="ybhb"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
MG1655: 80773"
residues 1 to 158 of 158 from Escherichia coli K-12 strain
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AG55144.1"
/db_xref="GI:12513759"
/translacion="MKLISNDLRDCKLPHRVNFGKQYDODNISPLAWDPVAKIR
SYVYCYDDPAPGSGMHWVNVNLPATRVLPQSGSLVAMPDGVLOTPTDQKIU
YCOAPFKGETHRITFVHALDVERIDVDEGASGAMGVNFHSLASITAMES"
complement(8203..9492)
/gene="bioA"
/note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gacattgtcgaagtcacagaatta 25
|||||
Db 9599 GACATTGTGCGACAGTCACAGANTTA 9623

RESULT 14
AP002553 297816 bp DNA BCF 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
DEFINITION AP002553 BA000007
ACCESSION AP002553.1 GI:13360211
VERSION
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuono,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak

JOURNAL
MEDLINE
REFERENCE
AUTHORS
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of RNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 297816)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-Info.osaka-u.ac.jp,
URL:ftp://www.gen-Info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.
FEATURES
Location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
/note="ECs0753"
/note="similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence="not experimental"
/product="sucinyl-coA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translacion="MNLHEYOAKOLFARYGLPAPVGYACTPREAEEASKIGAPPW
VKCVHAGGRGAGVYKVSIEDIRAFEMNMLGKLVLYOTDANGOPNOTIVPAAT
DIKELIYGAVDRSSRRVEMASTGVEITRKVAPPHILHKVALDPITGPMYOG
RELAEIKLEGKLVVOOFTKIPKGLATIFERLDALIEINPLVITKOGDILCDGKIGA
DGNALFROPDLREMRDSEDEPREAOAAWELINVAADENIGCVNAGAGLGMIDIV
AVAEVGVNVVVRLEGNAELGAKRIADSGINTIARAGLTDAAOVAAVEGK"
1245..2114
/gene="ECs0754"
1245..2114
/gene="ECs0754"
/note="ECs0754"
/note="similar to SUCC_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1

```


source

1. .5526
/organism="uncultured bacterium pCosH2"
/db_xref="taxon:143797"
/clone="pCosH2"
/note="unknown organism, cosmid clone derived from environmental consortium"
complement(52. .528)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kDa protein in modC-bioA intergenic region"
/protein_id="AAC60577.1"
/db_xref="GI:12620125"
/translation="MKRLSNDLPDGRKLPRIHVFNGMGYDGNISFHLAMDVPACTSEVYTCSDPDPAPTGSGMMHWVNLPAIDRVLPQGGSLVAMPGVQOTRTDFKACVDAAPKRGETHRIIFELVALDIERIDVDGASGAMVGFNVHFHLSASTAMS"
complement(587. .1876)
/gene="bioA"
complement(587. .1876)
/gene="bioA"
/note="7,8-diaminopelargonic acid synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase BioA"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDDLAFDQRIHMPYNSMTPLEVVVSABEGCELISDGSKRLVYGMSSMAWAIHGYNHPQLNAAMSQIDAMSHVFGI7HAPAIELCRKLVANTPPRLKCVPLADSSVAVEVAMKALQYMAKQAKRQFLFRNGYHGPTFGAMSCDPDONSMSHLMGYLEENLFAPAPROSGEDENWGFARLKAARHETLAATITPIYQAGGMSHMVPEWKLIRKICDREGILIDELATIGFTGKTFCEPAEILAPDILCLGRKLTGCTMTLSATLTREVALETINSGEAGCFMGTTFMGNLACANASIALISESDWQLOVADIEVOLREOLAPRADAENWADVRCALIGVETTHPVMAALQKFEFQGVMIIPFGLIILMPYITILPQOLIORLITAAVNAVQDETFFQ"
1963. .3003
/gene="bioB"
1963. .3003
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase bioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRPMTLSQVTELEFELDLLEAQQVNRHPPRQVOYSTLSTITGACPEPCKYCPQSSRYKIGLEAERLMEVEYVLEAARAKAAGSTFPGMAVKNIPNEDRIEOMYQVKGDLGECMTQIGLISQOORLANGLDYTNINLTSEFYVITNMLVAVKSGPLANDNDVDAFPFIRITAAIRIMPTSYVRSAGEEONNEQTQAMCFMAGNSLFYEGSKLLTTPNEPDKDLQFLRLGIPQOAVLAGDNEQOORLEQALMTPTDTRYAAAL"
3000. .4154
/gene="bioC"
3000. .4154
/gene="bioC"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase bioC"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSMOEKINALDNRARAALARRIYVAAGAGBMLVADROYLNPSSNYLISHPOTIRAWOGAGDFEIGSGSGSHVGSVYHQALEELAEMLGYSRALLEIFSGRANQAVIAAMMAEDRIADRLSHSLAEASLSPQLRFAHNVDYTLARLLASPCGGQQLVTEGVSMGDGSAPLAEITQVTOQNGMLWVDHAGTGVYGGEGKCSCHLQAKPELLVTFEKGSGVGAVALSSYVADYLLOFARHLIYSTSMRPAQOALRASLVARIRBEDQARREKGLSVLIAPRAGQDLPRLADSCAIDPLITGVNDSRALQIAEKIRQOQCVWTVAIRPTPYAGTARLRLITTAANEKQDIDRLLELVHNG"
4141. .4896
/gene="bioc"
4141. .4896
/gene="bioc"

```

/feature="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein BIOC"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVVKQA IAAFGRAALAEONADLQROSAVDVLTAMLPKRYT
HYLDAGCGGPMWT RHMREHAQVLTLDSPMLVOARQKADADHYLAGDI ESFPLATA
TFPLASNLNVAOCGNLSPALBELY RVRYSRGVYAFTTLYOGSLPELHQAQVADERP
HARFLPDEIEOSINGVYOHIDLPITNFDNALSNRSLKGIATGHTLHEGRPRLT
TSQQLRLDPEIAPQOQGRPLTHFLFVLAIRE"
4883 .>5526
/gene="biob"
4883 .>5526
/gene="biob"
/feature="PTB synthetase"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase BIOD"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRYFTVTGTEVCKTVAASCALLOAKAAGYRTAGKPYAS
GSEKPEGLRNSDALQIRNSQLDQATVNPATFAPTSPHLISAQGRPTSSMS
SIALAEQOADMVLVEGAGCFPLDPTFPFAPWYQEOPLVILVGVKLGCTINHAML
TAOLIOHGLTLAGVANDVTPKKRRAEITITLTKRTPALFLGELPWLAEENE"

BASE COUNT      1274 a      1507 c      1567 g      1178 t
ORIGIN
Query Match      84.0%; Score 21; DB 1; Length 5526;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      5 ttgtcgcaagtccagaatta 25
Db      1987 TTGTGCCAAGTCACAGAAATTA 2007
|||||
RESULT 16
AF248314
LOCUS      8227 bp      DNA      BCT      24-JAN-2001
DEFINITION      Uncultured bacterium pCosAS1 urocanase-like protein (nutH) gene,
partial cds: histidine ammonia-lyase-like protein (nutH), DAPA
amino transferase BIOC (biob), biotin synthase Biob (biob), 7-KAPA
synthetase (biob), biotin biosynthesis BIOC-like protein (bioc),
and dethiobiotin synthase Biob (biob) genes, complete cds: ABC
transporter-like protein (elSA) gene, partial cds; and unknown
gene.
AF248314
AF248314.1 GI:12407610
.
uncultured bacterium pCosAS1.
uncultured bacterium pCosAS1.
Bacteria: environmental samples.
1 (bases 1 to 8227)
Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streitel,W.R.
Direct Cloning from Enrichment Cultures, a Reliable Strategy for
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
2 (bases 1 to 8227)
Entcheva,P., Liebl,W. and Streitel,W.R.
Direct Submision
Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
Location/Qualifiers
1..8227
/organism="uncultured bacterium pCosAS1"
/db_xref="taxon:143793"
/clone="pCosAS1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
<1..502

```

```

gene
/ gene="hutu"
/ <1..502
/ gene="hutu"
/ note="putative urocanate hydratase"
/ codon_start=2
/ transl_table=11
/ product="urocanase-like protein"
/ protein_id="AAG53593.1"
/ db_xref="GI:12407618"
/ translation="PRERISFQGLTARICWGLEMRKLGAFNENVRGEYADTVI
GROHDSGASVSPNRETTEAMODSDAVSPMLNALTASGATWVSLHHGGVGMGF
SOHAGMYIVCDGTDAAARTARVLAHNDPATGVRRADAGYDIALIDCAKQGLNPMIA
ATQCOH"
502..2031
/ gene="huth"
502..2031
/ gene="huth"
/ note="putative"
/ codon_start=1
/ transl_table=11
/ product="histidine ammonia-lyase-like protein"
/ protein_id="AAG53586.1"
/ db_xref="GI:12407611"
/ translation="MMALTLTPTGLTTLAQLROYMOQPIQLTDESAREALINSVACVE
ATVAGRTAGINTGFLAQTATLHNDLENLRSYLSHAAGVGPPLDDIYRLMMV
LKIINSLARGESIRLVTOALIALVNGVTSVPAGSGASGDLAPLHMSTLLEGE
KARYRGEMLPAAATLQKAGLAEVTLAEGALLMCTQASTAFALRGLPEADLIFAS
AVCGALITTEAVLIGSRPPDARIHEVREGOQIDAAALFRHVLTDRSALESNCDK
VDDYSLRCOPQVQACITONROYAEVLIVESNADNPLVFAENEMVRFNFAEP
VMAADNIALAIEIGALSERRIALAMDKHMSPPLVNGVSGNFMIAQVTAAL
ASENKGICHPHTTSVDPKPCPSANDEHVSMAFAGRLMEAGNTRGVLAEMLAAL
QCADRDGLTSSPLLEDRKUSGGOVAHYDDRFAPDITAAISLTKSGLVLLPAC
L"
CDS
/ complement(2130..2606)
/ note="ORF1: similar to hypothetical 17.1 kDa protein in
mod-bioa intergenic region"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAG53587.1"
/ db_xref="GI:12407612"
/ translation="MKLISODLRDGDKLPLRHVNGMGYEGDNISPLAMDEYSGTK
SPVVCYDPDAPTOGSMWHIVANIADTRVLPGSGSSLYALPEGAVERTFPGAG
YGGAAPPKGERTIRYIFTHALNVERIEVDEEASAMGFNVHFTLGSASITAMYS"
/ complement(2651..3940)
/ gene="bioa"
/ complement(2651..3940)
/ note="7.8-diamino-pelargonic acid aminotransferase"
/ gene="bioa"
/ codon_start=1
/ transl_table=11
/ product="DAP aminotransferase BioA"
/ protein_id="AAG53588.1"
/ db_xref="GI:12407613"
/ translation="MTTDLAFDRHOIHMPYTSMTOPLPVYVVEAHGCELLASGEQ
LYDGSSWMAATHGVNHPDLAAMKAAQIDRMSHWEGITHOPAVDLCRLVATPPT
LTCVFPADSGSAVEYAKMMLQTYHAKGESRQRTTTRNGYHGDFTGAMCPDMS
MHSIMKGYLPENLFAPOSRFDGAMDEMDVGLALMAHRHEIAATLEPIYQAG
GRMYHPEWLKIRRCORECILLIADLAATGCFRTGLFACGAGITPDLICGKAL
TGTMTASATLTTROVAETISNGEAGCEFMHGTPEMGNPLACAVASESLALLESEMD
OYAAIESOLOEEELAVARDESEFADRYALGCVETTPVMAALQRFVYRGVWVP
FKLLIYMPYVIRPDQLSKLTRAVVDVAQVAHHTR"
4029..5069
/ gene="bioB"
4029..5069
/ gene="bioB"
/ gene="bioB"
/ codon_start=1
/ transl_table=11
/ product="biotin synthase BioB"
/ protein_id="AAG53589.1"
/ db_xref="GI:12407614"
/ translation="MAHQAARWTLQVYTALEKPLLELFEACQIHRQHPDQIQVST
LTSIKTGACPEDCKYCPOSARYKTGLESERLMEVQVLESARQANAGSTFCMGAAW

```

```

gene
/ gene="hutu"
/ <1..502
/ gene="hutu"
/ note="putative urocanate hydratase"
/ codon_start=2
/ transl_table=11
/ product="urocanase-like protein"
/ protein_id="AAG53593.1"
/ db_xref="GI:12407618"
/ translation="PRERISFQGLTARICWGLEMRKLGAFNENVRGEYADTVI
GROHDSGASVSPNRETTEAMODSDAVSPMLNALTASGATWVSLHHGGVGMGF
SOHAGMYIVCDGTDAAARTARVLAHNDPATGVRRADAGYDIALIDCAKQGLNPMIA
ATQCOH"
502..2031
/ gene="huth"
502..2031
/ gene="huth"
/ note="putative"
/ codon_start=1
/ transl_table=11
/ product="histidine ammonia-lyase-like protein"
/ protein_id="AAG53586.1"
/ db_xref="GI:12407611"
/ translation="MMALTLTPTGLTTLAQLROYMOQPIQLTDESAREALINSVACVE
ATVAGRTAGINTGFLAQTATLHNDLENLRSYLSHAAGVGPPLDDIYRLMMV
LKIINSLARGESIRLVTOALIALVNGVTSVPAGSGASGDLAPLHMSTLLEGE
KARYRGEMLPAAATLQKAGLAEVTLAEGALLMCTQASTAFALRGLPEADLIFAS
AVCGALITTEAVLIGSRPPDARIHEVREGOQIDAAALFRHVLTDRSALESNCDK
VDDYSLRCOPQVQACITONROYAEVLIVESNADNPLVFAENEMVRFNFAEP
VMAADNIALAIEIGALSERRIALAMDKHMSPPLVNGVSGNFMIAQVTAAL
ASENKGICHPHTTSVDPKPCPSANDEHVSMAFAGRLMEAGNTRGVLAEMLAAL
QCADRDGLTSSPLLEDRKUSGGOVAHYDDRFAPDITAAISLTKSGLVLLPAC
L"
CDS
/ complement(2130..2606)
/ note="ORF1: similar to hypothetical 17.1 kDa protein in
mod-bioa intergenic region"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAG53587.1"
/ db_xref="GI:12407612"
/ translation="MKLISODLRDGDKLPLRHVNGMGYEGDNISPLAMDEYSGTK
SPVVCYDPDAPTOGSMWHIVANIADTRVLPGSGSSLYALPEGAVERTFPGAG
YGGAAPPKGERTIRYIFTHALNVERIEVDEEASAMGFNVHFTLGSASITAMYS"
/ complement(2651..3940)
/ gene="bioa"
/ complement(2651..3940)
/ note="7.8-diamino-pelargonic acid aminotransferase"
/ gene="bioa"
/ codon_start=1
/ transl_table=11
/ product="DAP aminotransferase BioA"
/ protein_id="AAG53588.1"
/ db_xref="GI:12407613"
/ translation="MTTDLAFDRHOIHMPYTSMTOPLPVYVVEAHGCELLASGEQ
LYDGSSWMAATHGVNHPDLAAMKAAQIDRMSHWEGITHOPAVDLCRLVATPPT
LTCVFPADSGSAVEYAKMMLQTYHAKGESRQRTTTRNGYHGDFTGAMCPDMS
MHSIMKGYLPENLFAPOSRFDGAMDEMDVGLALMAHRHEIAATLEPIYQAG
GRMYHPEWLKIRRCORECILLIADLAATGCFRTGLFACGAGITPDLICGKAL
TGTMTASATLTTROVAETISNGEAGCEFMHGTPEMGNPLACAVASESLALLESEMD
OYAAIESOLOEEELAVARDESEFADRYALGCVETTPVMAALQRFVYRGVWVP
FKLLIYMPYVIRPDQLSKLTRAVVDVAQVAHHTR"
4029..5069
/ gene="bioB"
4029..5069
/ gene="bioB"
/ gene="bioB"
/ codon_start=1
/ transl_table=11
/ product="biotin synthase BioB"
/ protein_id="AAG53589.1"
/ db_xref="GI:12407614"
/ translation="MAHQAARWTLQVYTALEKPLLELFEACQIHRQHPDQIQVST
LTSIKTGACPEDCKYCPOSARYKTGLESERLMEVQVLESARQANAGSTFCMGAAW

```

```

BASE COUNT      1512 a      2642 c      2605 g      1453 t      15 others
ORIGIN
Query Match      72.0%; Score 18; DB 1; Length 8227;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1      gacattgtcgaagtcac 18
Db      4049 GACATTGTGCAAGTCAC 4066

RESULT  17
AC087305/c
LOCUS   AC087305.1
DEFINITION Homo sapiens chromosome 18 clone RP11-238P13 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC087305
VERSION   AC087305.1 GI:11990696
KEYWORDS  HTG; HTGS_PHASE0.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS  Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Homo sapiens chromosome 18, clone RP11-238P13
JOURNAL   Unpublished
REFERENCE
AUTHORS  2 (bases 1 to 32767)
          Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
          Barina,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
          Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
          Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
          Dodge,S., Faro,S., Ferrelia,P., Fitzhugh,W., Gage,D., Galagan,J.,
          Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
          Jones,C., Karatas,A., Laroque,K., Lamares,R., Landers,T.,
          Lebeck,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
          Marquis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
          McPheters,R., Meldrum,J., Menus,L., Minova,T., Mlenga,V.,
          Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
          O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
          Plunkhang,P., Pierre,N., Pollata,V., Raymond,C., Retta,R.,
          Riebeck,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M.,
          Rouey,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
          Sougniez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
          Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
          Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zaitoun,J.,
          Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: l12071
Center clone name: 238_P_13

* NOTE: This record contains 40 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1       707: contig of 707 bp in length
*       708 807: gap of 100 bp
*       808 1527: contig of 720 bp in length
*       1528 1627: gap of 100 bp

1628      2348: contig of 721 bp in length
2349      2448: gap of 100 bp
2449      3157: contig of 709 bp in length
3158      3257: gap of 100 bp
3258      3973: contig of 716 bp in length
3974      4073: gap of 100 bp
4074      4799: contig of 726 bp in length
4800      4899: gap of 100 bp
4900      5617: contig of 718 bp in length
5618      5717: gap of 100 bp
5718      6450: contig of 733 bp in length
6451      6550: gap of 100 bp
6551      7259: contig of 709 bp in length
7260      7359: gap of 100 bp
7360      8066: contig of 707 bp in length
8067      8166: gap of 100 bp
8167      8907: contig of 741 bp in length
8908      9007: gap of 100 bp
9008      9716: contig of 709 bp in length
9717      9816: gap of 100 bp
9817      10532: contig of 716 bp in length
10533      10632: gap of 100 bp
10633      11354: contig of 722 bp in length
11355      11454: gap of 100 bp
11455      12166: contig of 712 bp in length
12167      12266: gap of 100 bp
12267      12976: contig of 710 bp in length
12977      13076: gap of 100 bp
13077      13789: contig of 713 bp in length
13790      13889: gap of 100 bp
13890      14617: contig of 728 bp in length
14618      14717: gap of 100 bp
14718      15452: contig of 735 bp in length
15453      15552: gap of 100 bp
15553      16276: contig of 724 bp in length
16277      16376: gap of 100 bp
16377      17111: contig of 735 bp in length
17112      17211: gap of 100 bp
17212      17925: contig of 714 bp in length
17926      18025: gap of 100 bp
18026      18746: contig of 721 bp in length
18747      18846: gap of 100 bp
18847      19569: contig of 723 bp in length
19570      19669: gap of 100 bp
19670      20384: contig of 715 bp in length
20385      20484: gap of 100 bp
20485      21190: contig of 706 bp in length
21191      21290: gap of 100 bp
21291      22028: contig of 738 bp in length
22029      22128: gap of 100 bp
22129      22841: contig of 713 bp in length
22842      22941: gap of 100 bp
22942      23653: contig of 712 bp in length
23654      23753: gap of 100 bp
23754      24478: contig of 725 bp in length
24479      24578: gap of 100 bp
24579      25305: contig of 727 bp in length
25306      25405: gap of 100 bp
25406      26143: contig of 738 bp in length
26144      26243: gap of 100 bp
26244      26968: contig of 725 bp in length
26969      27068: gap of 100 bp
27069      27785: contig of 717 bp in length
27786      27885: gap of 100 bp
27886      28584: contig of 699 bp in length
28585      28684: gap of 100 bp
28685      29467: contig of 783 bp in length
29468      29567: gap of 100 bp
29568      30292: contig of 725 bp in length
30293      30392: gap of 100 bp
30393      31117: contig of 725 bp in length
31118      31217: gap of 100 bp
31218      31944: contig of 727 bp in length

```

```

* 31945 32044: gap of 100 bp
* 32045 32767: contig of 723 bp in length.
Location/Qualifiers
1. 32767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-238P13"
/clone_lib="RPCT-11 Human Male BAC"
BASE COUNT 8045 a 5670 c 5396 g 9330 t 4326 others
ORIGIN
Query Match 64.0%; Score 16; DB 2; Length 32767;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaatta 25
|||||
Db 30730 GCAAGTCACAGAATTA 30715

RESULT 18
AC021110 155172 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 18 clone RP11-238P13, WORKING DRAFT
DEFINITION SEQUENCE, 6 unordered pieces.
ACCESSION AC021110.3 GI:8439978
VERSION AC021110.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 155172)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155172)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 10, 2000 this sequence version replaced gi:7021826.

COMMENT ----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information -----
Center project name: H_NH0238P13
Summary Statistics -----
Sequencing vector: Plasmid; 73%
Chemistry: Dye-primer ET; 73% of reads
Chemistry: Dye-terminator Big Dye; 27% of reads
Assembly program: Phrap; version 0.960319
Consensus quality: 151162 bases at least Q40
Consensus quality: 152443 bases at least Q30
Consensus quality: 153249 bases at least Q20
Insert size: 177000; agarose-efp
Insert size: 154672; sum-of-contigs
Quality coverage: 4.08 in Q20 bases; sum-of-contigs
Quality coverage: 4.69 in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

```

* as soon as it is available and the accession number will
* be preserved
1
* 9265: contig of 9265 bp in length
* 9266 9365: gap of unknown length
* 9366 26504: contig of 17139 bp in length
* 26505 26604: gap of unknown length
* 26605 44300: contig of 17696 bp in length
* 44301 44400: gap of unknown length
* 44401 62250: contig of 17850 bp in length
* 62251 62350: gap of unknown length
* 62351 105744: contig of 43394 bp in length
* 105745 105844: gap of unknown length
* 105845 155172: contig of 49328 bp in length.
Location/Qualifiers
1. 155172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/clone="RP11-238P13"
1. 9265
/clone="assembly_name:Contig1"
9366 26504
/clone="assembly_name:Contig2"
26605 44300
/clone="assembly_name:Contig3"
clone_end:77
vector_side:right"
44401 62250
/clone="assembly_name:Contig4"
62351 105744
/clone="assembly_name:Contig5"
105845 155172
/clone="assembly_name:Contig6"
BASE COUNT 51543 a 28385 c 28043 g 46698 t 503 others
ORIGIN
Query Match 64.0%; Score 16; DB 2; Length 155172;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaatta 25
|||||
Db 111929 GCAAGTCACAGAATTA 111944

RESULT 19
AC041018/c 156184 bp DNA HTG 22-MAY-2000
LOCUS Homo sapiens chromosome 4 clone RP11-478020 map 4, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
ACCESSION AC041018.2 GI:8016453
VERSION AC041018.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 156184)
AUTHORS Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barina,N., Bastien,V., Bede,F.,
Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Labrecque,K., Lamazares,R., Landers,T., Lehocaky,J.,

```

TITLE
JOURNAL
COMMENT

Levine, R., Liew, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McKus, A., McKernan, K., McPheeters, R., Melidrin, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teledge, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 22, 2000 this sequence version replaced g1:7534220.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1597)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center Project name: L9577
Center clone name: 478_O_20

Summary Statistics

Sequencing vector: M13: M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 147958 bases at least Q40
Consensus quality: 151874 bases at least Q30
Consensus quality: 153413 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 154684; sum-of-ctrls

Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      1351: contig of 1351 bp in length
*      1352 1451: gap of 100 bp
*      1452 4369: contig of 2918 bp in length
*      4370 4469: gap of 100 bp
*      4470 8873: contig of 4404 bp in length
*      8874 8973: gap of 100 bp
*      8974 13167: contig of 4194 bp in length
*      13168 13267: gap of 100 bp
*      13268 18889: contig of 5722 bp in length
*      18890 19089: gap of 100 bp
*      19090 24989: contig of 5900 bp in length
*      24990 25089: gap of 100 bp
*      25090 33460: contig of 8371 bp in length
*      33461 33560: gap of 100 bp
*      33561 42328: contig of 8768 bp in length
*      42329 42428: gap of 100 bp
*      42429 50674: contig of 8246 bp in length
*      50675 50774: gap of 100 bp
*      50775 59862: contig of 9088 bp in length
*      59863 59962: gap of 100 bp
*      59963 71799: contig of 11737 bp in length
*      71700 71799: gap of 100 bp
*      71800 82583: contig of 10784 bp in length
*      82584 82683: gap of 100 bp
*      82684 95072: contig of 12389 bp in length
*      95073 95172: gap of 100 bp
*      95173 108949: contig of 13777 bp in length

```

```

*      108950 109049: gap of 100 bp
*      109050 126668: contig of 17619 bp in length
*      126669 126768: gap of 100 bp
*      126769 156184: contig of 29416 bp in length.
Location/Qualifiers
1. 156184
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-478020"
/clone_lib="RP11 Human Male BAC"
1. 1351
/note="assembly-fragment"
1452. 4369
/note="assembly-fragment"
4470. 8873
/note="assembly-fragment"
clone_end:SP6
vector_side:right"
8974. 13167
/note="assembly-fragment"
13268. 18889
/note="assembly-fragment"
19090. 24989
/note="assembly-fragment"
25090. 33460
/note="assembly-fragment"
33561. 42328
/note="assembly-fragment"
42429. 50674
/note="assembly-fragment"
50775. 59862
/note="assembly-fragment"
59963. 71699
/note="assembly-fragment"
71800. 82583
/note="assembly-fragment"
82684. 95072
/note="assembly-fragment"
clone_end:T7
vector_side:right"
95173. 108949
/note="assembly-fragment"
109050. 126668
/note="assembly-fragment"
126769. 156184
/note="assembly-fragment"

```

```

BASE COUNT 51507 a 26941 c 26653 g 49578 t 1505 others
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 2; Length 156184;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 gcaagtcacagaatta 25
|||||
Db 123532 GCAAGTCACAGAATTA 123517

```

```

RESULT 20
AP001390/c 174815 bp DNA HTG 30-MAY-2000
LOCUS Homo sapiens chromosome 18 clone RP11-655B11 map 18q23, WORKING
DEFINITION DRAFT SEQUENCE. 21 unordered pieces.
ACCESSION AP001390
VERSION AP001390.2 GI:8117301
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-655B11.
ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

```



```

misc_feature      /note="assembly_fragment"
159294..161359
/note="assembly_fragment clone_end:v7 vector_side:right"
misc_feature      /note="assembly_fragment clone_end:SP6 vector_side:left"
161460..163515
/note="assembly_fragment"
163616..166614
/note="assembly_fragment"
166715..169056
/note="assembly_fragment"
169157..171527
/note="assembly_fragment"
171628..173248
/note="assembly_fragment"
173349..174815
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
52310 a 30923 c 32183 g 57399 t 2000 others
BASE COUNT
ORIGIN

```

```

Query Match      64.0%: Score 16; DB 2; Length 174815;
Best Local Similarity 100.0%: Pred. No. 5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 gcaagtcacagaatta 25
|||||
Db 44384 GCAAGTCACAGAAATTA 44369

```

```

RESULT 21
AC091589/c      AC091589 178056 bp DNA HTG 26-JUL-2001
LOCUS Homo sapiens chromosome 18 clone RP11-635B11 map 18, *** SEQUENCING
DEFINITION
AC091589
IN PROGRESS ***, 2 ordered pieces.
AC091589
HTG: HTGS-PHASE2; HTGS-FULLTOP; HTGS-ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
TITLE Homo sapiens chromosome 18, clone RP11-635B11
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 178056)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
TITLE Homo sapiens chromosome 18, clone RP11-635B11
JOURNAL Unpublished
AUTHORS

```

```

TITLE
JOURNAL
COMMENT
Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 26, 2001 this sequence version replaced g1:14994291.
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web Site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L13182
Center clone name: 635_B_11
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 168134: contig of 168134 bp in length
* 168135 168234: gap of 100 bp
* 168235 178056: contig of 9822 bp in length.
Location/Qualifiers
1. 178056
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-635B11"
/clone_lib="RP11-635B11 Human Male BAC"
BASE COUNT 49585 a 30884 c 34094 g 63393 t 100 others
ORIGIN

```

```

Query Match      64.0%: Score 16; DB 2; Length 178056;
Best Local Similarity 100.0%: Pred. No. 5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 gcaagtcacagaatta 25
|||||
Db 64175 GCAAGTCACAGAAATTA 64160

```

```

RESULT 22
AC020734/c      AC020734 196550 bp DNA PRI 09-AUG-2001
LOCUS Homo sapiens chromosome 4 clone RP11-655114, complete sequence.
DEFINITION
AC020734
AC020734
HTG: HTGS-PHASE2; HTGS-FULLTOP; HTGS-ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 196550)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 196550)
REFERENCE Waterston,R.H.
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 9, 2001 this sequence version replaced g1:15022773.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center

```


Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@atson.wustl.edu
----- Project Information -----
Center project name: H_NH065114

FEATURES
source
Location/Qualifiers
1. 196550
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-665114"

BASE COUNT 65000 a 34953 c 34110 g 62487 t
ORIGIN

Query Match 64.0%; Score 16; DB 9; Length 196550;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatta 25
|||||
DB 55717 GCAAGTCACAGATT 55702

RESULT 23

AX136747 561 bp DNA PAT 30-MAY-2001
LOCUS AX136747
DEFINITION Sequence 669 from Patent EP1067182.
ACCESSION AX136747
VERSION AX136747.1 GI:14273151
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 561)
Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
Hayashi,K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 669 10-JAN-2001;
Helix Research Institute (JP)
FEATURES
source
1. 561
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 213 a 83 c 114 g 149 t 2 others
ORIGIN

Query Match 60.0%; Score 15; DB 6; Length 561;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
|||||
DB 339 CAAAGTCACAGATT 353

RESULT 24
AR080743 1771 bp DNA PAT 31-AUG-2000
LOCUS AR080743
DEFINITION Sequence 48 from patent US 5968826.
ACCESSION AR080743
VERSION AR080743.1 GI:10007473
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
1 (bases 1 to 1771)
Bennett,C.,Frank,C.,Condon,T.P. and Cowser,L.M.
TITLE Antisense inhibition of integrin alpha4 expression
JOURNAL Patent: US 5968826-A 48 19-OCT-1999;

FEATURES
source
Location/Qualifiers
1. 1771
/organism="unknown"
BASE COUNT 371 a 471 c 519 g 410 t
ORIGIN

Query Match 60.0%; Score 15; DB 6; Length 1771;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatta 24
|||||
DB 312 GCAAGTCACAGATT 326

RESULT 25
MUSINTESUB 1771 bp DNA ROD 18-APR-1996
LOCUS MUSINTESUB
DEFINITION Mus musculus (clone AAG16) Integrin alpha-4 subunit gene, 5' end
cds
L20788
VERSION L20788.1 GI:309416
KEYWORDS integrin alpha-4 subunit.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1771)
De Meirman,C., Scholten,E., Jaepers,M., Ongena,K., Matthijs,G.,
Marynen,P. and Cassiman,J.J.
TITLE Cloning and characterization of the promoter region of the murine
alpha-4 integrin subunit
JOURNAL DNA Cell Biol. 13 (7), 743-754 (1994)
MEDLINE 95290094
FEATURES
source
Location/Qualifiers
1. 1771
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="fibroblasts"
/cell_line="3T3"
/tissue="fibroblasts"
/note="putative"
503..508
/note="putative"
845..850
952..1389
complement(1055..1060)
join(1193..1389,1711..1771)
/gene="VLA-4"
join(1193..1389,1711..>1771)
/gene="VLA-4"
/note="putative"
/codon_start=1
/product="integrin alpha-4 subunit"
/protein_id="AAA97501.1"
/db_xref="GI:309417"
/translation="MAAEARCRPRSRGIALREAVMLLYGCVPTGHSTNDPENALLY
CGPSGTLFGYSVVLVHSHGSKRWLVIGAPTASMLSNASVNPBG"
1390..1710
1390..1771
1711..1771
/gene="VLA-4"

BASE COUNT 371 a 471 c 519 g 410 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 1771;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatta 24
|||||
DB 312 GCAAGTCACAGATT 326

RESULT 26
 LOCUS G07181/c 1329 bp DNA STS 19-OCT-1995
 DEFINITION human STS WI-9211.
 ACCESSION G07181
 VERSION G07181.1 GI:860426
 KEYWORDS STS sequence; primer; sequence tagged site.
 SOURCE human STS derived from sequences in dbest and the Unigene collection.
 ORGANISM Homo sapiens
 Eukaryota; Eukaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 1829)
 AUTHORS Hudson, T.
 TITLE Whitehead Institute/MIT Center for Genome Research: Physically Mapped ESTs
 JOURNAL Unpublished (1995)
 COMMENT Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu
 Primer A: CAGCCACACTACTGCAATTAGC
 Primer B: GACAGAAATGATTTGTATTAATGCG
 STS size: 111
 PCR profile:
 Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:
 Protocol:
 Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul
 Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3
 FEATURES Prepared with primer pairs derived from D13644 -- Unigene.
 source Location/Qualifiers
 1..1829
 STS /organism="Homo sapiens"
 primer_bind 891..1001
 primer_bind 891..912
 complement(977..1001)
 BASE COUNT 539 a 321 c 320 g 590 t 59 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; length 1829;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caagtcacagacta 25
 |||||||||||||
 DB 1689 CAAGTCACAGACTTA 1675
 RESULT 27

MMINTGRO1
 LOCUS MMINTGRO1 1949 bp DNA ROD 02-OCT-1996
 DEFINITION Mus musculus alpha 4 integrin gene, exons 1 and 2.
 ACCESSION U34627
 VERSION U34627.1 GI:1173582
 KEYWORDS 1 of 22
 SEGMENT house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS De Meirman, C., Scholten, E., Jaspers, M., Ongena, K., Mathijs, G., Marynen, P., and Cassiman, J.J.
 TITLE Cloning and characterization of the promoter region of the murine alpha-4 integrin subunit
 JOURNAL DNA Cell Biol. 13 (7), 743-754 (1994)
 MEDLINE 95290094
 REFERENCE 2 (bases 1 to 1949)
 AUTHORS De Meirman, C., Jaspers, M., Scholten, E., and Cassiman, J.J.
 TITLE The genomic structure of the murine alpha 4 integrin gene
 JOURNAL DNA Cell Biol. 15 (7), 595-603 (1996)
 MEDLINE 96326295
 REFERENCE 3 (bases 1 to 1949)
 AUTHORS De Meirman, C.
 TITLE Direct Submission
 JOURNAL Submitted (23-AUG-1995) Catherine De Meirman, Center for Human Genetics, University of Leuven, Herestraat 49, Leuven 3000, Belgium
 FEATURES location/Qualifiers
 source 1..1949
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /tissue_type="spleen"
 /note="NIH-3T3 B6/CBA spleen genomic library in lambda fix
 fixit"
 II: NIH-3T3 nd B6/CBA spleen genomic library in lambda
 fixit"
 promoter 1..952
 exon 953..1389
 /number=1
 5'UTR 953..1192
 exon 1711..1832
 /number=2
 BASE COUNT 420 a 514 c 563 g 452 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 10; length 1949;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 gcaagtcacagact 24
 |||||||||||||
 DB 312 GCAGTCACAGACTT 326
 RESULT 28
 ECHK311M
 LOCUS ECHK311M 2340 bp DNA BCT 25-SEP-1998
 DEFINITION E. coli HK311M gene.
 ACCESSION X82231
 VERSION X82231.1 GI:3660492
 KEYWORDS alpha-polypeptide; beta-polypeptide; R.ECHK311 protein.
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 2340)
 AUTHORS Lee, K.F., Kam, K.M., and Shaw, P.C.
 TITLE A bacterial methyltransferase M.ECHK311 requires two proteins for in vitro methylation
 JOURNAL Nucleic Acids Res. 23 (1), 103-108 (1995)
 MEDLINE 95175351
 REFERENCE 2 (bases 1 to 2340)

AUTHORS Shaw, P.
 TITLE Direct Submission
 JOURNAL Submitted (18-OCT-1994) P. Shaw, Dept. of Biochemistry, The Chinese University of Hong Kong, Shatin NT, HONG KONG
 REMARK 3 (bases 1 to 2340)
 AUTHORS Shaw, P.
 TITLE Direct Submission
 JOURNAL Submitted (16-DEC-1997) P. Shaw, Dept. of Biochemistry, The Chinese University of Hong Kong, Shatin NT, HONG KONG
 REFERENCE 4 (bases 1 to 2340)
 Lee, K.F., Shaw, P.C., Picone, S.J., Wilson, G.G. and Lunnen, K.D.
 Sequence comparison of the EcoHK311 and EaeI restriction-modification systems suggests an intergenic transfer of genetic material
 JOURNAL Biol. Chem. 379 (4-5), 437-441 (1998)
 COMMENT On Sep 28, 1998 this sequence version replaced gi:639971.
 FEATURES
 SOURCE

1..2340
 /organism="Escherichia coli"
 /strain="HK31"
 /db_xref="taxon:562"
 407..1336
 /gene="ecoHK311M alpha"
 407..1336
 /gene="ecoHK311M alpha"
 /function="methyltransferase"
 /codon_start=1
 /transl_table=1
 /product="M.EcoHK311 alpha polypeptide"
 /protein_id="CA57706.1"
 /db_xref="GI:3660493"
 /db_xref="SPTREMBL:087740"
 /translation="MKKKPLKQYKVSFAGIGFDLGKAGHEVFOCEINRCKOK
 VLKNNKSVPLHDTIRLNADIEPESNVMWCCGPPCDVSSANGKKGLEGARSGFLY
 TYALKIERKPEMLIENVGLNSHNGDFKVIIDTLYVEFGVSMRYLDKAYECITP
 ORRRRYVASIGMRSARVLEFGARIVDQGGCFRAASGASATGLSKADIVSIQ
 HASTGRATAPQKQKRYRNDCEFTYTTIDSGSSNDVAVCPADAFVRVASTGTSRLDSNR
 FRVAGNVAVPIVEMLGRIVLDQVLEALS"
 717..1247
 /gene="ecoHK311M beta"
 717..1247
 /gene="ecoHK311M beta"
 /function="methyltransferase"
 /codon_start=1
 /transl_table=1
 /product="M.EcoHK311 polypeptide beta"
 /protein_id="CA57705.1"
 /db_xref="GI:3660494"
 /db_xref="SPTREMBL:047257"
 /translation="MNSSKRESLNGLLKMEPPCSTATMDTKTSLSSTNSMGAF
 LGAYWMNTLHPSPVEECITLSVLETCAPIESPLNEPQLSLINRAKEGOMLPEPL
 LQAYKOISILSMQVDEKOPDKOKDGTCTMKRPILTLQEEVOMLYVRMLPSEVE
 RLOGPENNWTLIDSSO"
 complement(1317..2273)
 /gene="ecoHK311R"
 complement(1317..2273)
 /gene="ecoHK311R"
 /function="restriction endonuclease"
 /codon_start=1
 /transl_table=1
 /product="R.EcoHK311 protein"
 /protein_id="CA57707.1"
 /db_xref="GI:3660495"
 /db_xref="SPTREMBL:087741"
 /translation="MAVSIAYELGQRTDQDIIVPAKGTQIQCFEMDRICDKASKVK
 NPVPCSVARKPDGTWIVCEHRLCSTROKRTVJNGKKQIENILVEHQRDILIRKVA
 KTIYDPELODESGRREVNILPDSNNSYADYVMKNSFGRCRVDEGOMLPEPL
 TSTSGEITHIAAMDLEFPNTNELLROVANTJETNAMRROGEOPLVKGVMYDQSG
 KIVFAVGLSDYLRFRNRNANLRLKHNNTLCLAFKEDTSDPRPCPLIPLIDS
 KTLFTNYSTFVRLTDGAPREPLEGSGFLKLDSSVTIMPR"

BASE COUNT

679 a 458 c 527 g 676 t

ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 2340;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 acatgtcgcgaagc 16
 Db 906 ACATGTCGCAGTC 920

RESULT 29
 LOCUS Y029487S18
 DEFINITION Homo sapiens aminophospholipid-transporting ATPase (ATP10C) gene,
 exons 20, 21, and complete cds.
 ACCESSION Y029504.1 GI:13878298
 VERSION Y029504
 KEYWORDS
 SEGMENT
 SOURCE
 ORGANISM
 human
 18 of 18
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 3140)
 Herzog, L.B.K., Kim, S.J., Cook, E.H. Jr. and Ledbetter, D.H.
 The human aminophospholipid-transporting ATPase gene ATP10C maps
 adjacent to UBE3A and exhibits similar imprinted expression
 Am. J. Hum. Genet. 68 (6), 1501-1505 (2001)

REFERENCE
 AUTHORS Herzog, L.B.K., Kim, S.-J., Cook, E.H. Jr. and Ledbetter, D.H.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-2001) Human Genetics and Psychiatry, University of Chicago, 920 E. 58th Street, Chicago, IL 60637, USA
 FEATURES
 SOURCE
 location/Qualifiers
 1..3140
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 join(AT029487.1:1..582,AY029488.1:96..300,
 AY029489.1:118..203,AY029490.1:1..107,AY029491.1:34..165,
 AY029492.1:713..843,AY029493.1:58..310,AY029494.1:88..305,
 AY029495.1:75..269,AY029496.1:102..669,
 AY029497.1:185..288,AY029497.1:383..509,
 AY029498.1:99..283,AY029499.1:88..415,AY029500.1:145..221,
 AY029501.1:119..244,AY029502.1:148..348,
 AY029503.1:177..257,AY029503.1:337..441,105..292,
 439..3140)
 /gene="ATP10C"
 /product="aminophospholipid-transporting ATPase"
 order(AY029487.1:1..1040,AY029488.1:1..331,
 AY029489.1:1..357,AY029490.1:1..107,AY029491.1:1..333,
 AY029492.1:1..960,AY029493.1:1..436,AY029494.1:1..438,
 AY029495.1:1..396,AY029496.1:1..743,AY029497.1:1..588,
 AY029498.1:1..421,AY029499.1:1..558,AY029500.1:1..327,
 AY029501.1:1..269,AY029502.1:1..459,AY029503.1:1..487,
 1..3140)
 /gene="ATP10C"
 join(AT029487.1:134..582,AY029488.1:96..300,
 AY029489.1:118..203,AY029490.1:1..107,AY029491.1:34..165,
 AY029492.1:713..843,AY029493.1:58..310,AY029494.1:88..305,
 AY029495.1:75..269,AY029496.1:102..669,
 AY029497.1:185..288,AY029497.1:383..509,
 AY029498.1:99..283,AY029499.1:88..415,AY029500.1:145..221,
 AY029501.1:119..244,AY029502.1:148..348,
 AY029503.1:177..257,AY029503.1:337..441,105..292,
 439..1072)
 /gene="ATP10C"
 /codon_start=1
 /product="aminophospholipid-transporting ATPase"
 /protein_id="AAK33100.1"

CDS

gene

mRNA

[illegible]

3' UTR	BASE COUNT	ORIGIN
1340 a	1041 c	995 g 1222 t
2767 .	4602	

	Query Match	60.0%	Score 15;	DB 9;	Length 4602;
	Best Local Similarity	100.0%;	Pred. No. 36;		
	Matches 15; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;
OY	11 caagtcacagaattca	25 			
Db	4462 CAAGTCACAGAATTAA	4448			
RESULT 33					
AF039582 LOCUS DEFINITION	AP039582 4636 bp DNA BCT 05-DEC-2000				
	Enterobacter aerogenes putative integrase (int(p)) gene, partial cds; and EaeI methyltransferase alpha subunit (eaeIM-a), EaeI methyltransferase beta subunit (eaeIM-b), and EaeI restriction endonuclease (eaeRI) genes, complete cds.				
ACCESSION VERSION KEYWORDS SOURCE ORGANISM	AF039582 AF039582.1 GI:2760953 Enterobacter aerogenes. Enterobacter aerogenes. Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae; Enterobacter. 1 (bases 1 to 4636) Lee,K.F., Shaw,P.C., Picone,S.J., Wilson,G.G. and Lunnen,K.D. Sequence comparison of the EcoHK31I and EaeI restriction-modification systems suggests an intergenic transfer of genetic material Biol.Chem. 379 (4-5), 437-441 (1998)				
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL FEATURES	Biol.Chem. 379 (4-5), 437-441 (1998) 98290239 9628335 2 (bases 1 to 4636) Lunnen,K.D., Picone,S.J. and Wilson,G.G. Direct Submission Submitted (22-DEC-1997) Research, New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA Location/Qualifiers 1..4636 /organism="Enterobacter aerogenes" /db_xref="taxon:548" 				
CDS	<1..1035 /gene="int(p)" >1..1035 /note="int(p)" /note="similar to phage P4 integrase SP:P9347"				
	/codon_start=1 /transl_table=1 /product="putative integrase"				
	/protein_id="BAB95339.1"				
	/db_xref="GI:2760957"				
	/translation="KLVVAGIDPRHKNVAEEDTKGATITPESVAREMHANKWPEEHSRLVALSLEDNLFPALIGKRISIDSFTRDLLVPILKEATGRLEVASRLOORTAHLIR YAVOGGLIDYNPAQMGAVASNNVRHPALELRIPPELLORIDSYSGRPETRLAVBL TLTITRSSELDFARWEPIDEDERAMWTTPAPEREAIEGVNHSRGSKMTPLVLSPQA ALFIKKOYHFKSEGERDVEFGVDHNFRKMPSNTYAKALKRVMGDYDTKYEVCGHGRTMAA GSLSIESGLMSDAVEROMSHMERSVAAYIIHKAHEIDERKLMIQLMWADLDANROK GISPFDIKA"				
gene	1128..2157 /gene="eaeIM-a"				
CDS	1128..2157 /gene="eaeIM-a"				
	/function="converts yggCCR to ygmcCCR"				
	/note="M.EaeIa"				
	/codon_start=1 /transl_table=1				
	/product="EaeI methyltransferase alpha subunit"				
	/protein_id="BAB95336.1"				
	/db_xref="GI:2760954"				
	/translation="MKKRPLKOYKVSSFAGIGCFDLGLEKAGMEVVQEENIKFQQ VLRKANPNYLPHMDITPOLNABDIRPSSVMGGSPCOOVASNOCKRKLFEGARGLFLFY ITAKLIERRKPMLVIENVPGILNSHNDOPFVVIDTVELCYGSNRYLDAKKFGTFE				

ORRRRYIVASLGDMNSARVLEPCATRVYKQGGERTNASCASDTGLSKADIVS1Q
 HASLGRKATGPAQKGYRNDGERTYTLTDSRSSDCAVCAPIAAFRVARSAGISRELDNR
 FRAVGNAAVAPITIEWIGKRIYLVDQOOLNLEALS"
 1538..2068
 /gene="eaeIM-b"
 1538..2068
 /gene="eaeIM-b"
 /note="M.EaeIb; translated from 3'-half of the eaeIM-a
 gene in an alternative reading frame"
 /codon_start=1
 /transl_table=11
 /product="EaeI methyltransferase beta subunit"
 /protein_id="AAB95337.1"
 /db_xref="GI:2760955"
 /translation="MONSSKRENLNGLLKPEPCSTATMDKTSKLSIRMSNGMAF
 RGRYMNLTLEHPSEECTLSOVLETCAPLESFLNPEQLINAKERGQMLPEPI
 IQAYKQOISILSSOVLDEKOPDLKOKDGTTEKPTHTLQEEVQMLVYRRMLPSEYE
 RLGPEENWTLIDSEQ"
 complement(2138..3085)
 /gene="eaeIR"
 complement(2138..3085)
 /gene="eaeIR"
 /function="Cleaves between Y and G in YGGCCR"
 /note="R.EaeI; unable to cleave methylated ds sequence
 YGGCCR"
 /codon_start=1
 /transl_table=11
 /product="EaeI restriction endonuclease"
 /protein_id="AAB95338.1"
 /db_xref="GI:2760956"
 /translation="MAVSAEYLGORTVDQIIVPSKGTIAQCFPMRTCDKSKNN
 HVCGRKRPDLTWTCEHRLCSTROKKTVISGERTQOQDNVLVEHODILRKAKLI
 YDPELODESEIGVKKREVNIPDSDNSYHADYVMNRSRGVDEVLLEMGCGCTSS
 TGAIRHIAAMADLEFPINELIRPVANLITETNARQOQDFLVKKVAVVOTGKIV
 FVNSLIDYIHRRRRNALRDLKANNMTICLAKEDTSDPEQGPILITDSTKL
 FVNSTFVAFRLTDCAGPPELPESSFLTIDNSTVTIMPR"

BASE COUNT 1362 a 893 c 1074 g 1307 t

ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 4636;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 aaatgtcgaatgc 16
 |||||||||||||
 Db 1727 ACATTGTCGCAAGTC 1741

RESULT 34
 AF152313 5218 bp mRNA PRI 22-JUL-1999
 LOCUS AF152313 Homo sapiens protocadherin alpha 5 (PCDH-alpha5) mRNA, complete
 DEFINITION cds.
 ACCESSION AF152313
 VERSION AF152313.1 GI:5456913
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE Wu.Q. and Maniatis,T.
 A striking organization of a large family of human neural
 cadherin-like cell adhesion genes
 JOURNAL Cell 97 (6), 779-790 (1999)
 MEDLINE 99308636
 PUBMED 10380929
 2 (bases 1 to 5218)
 Wu.Q. and Maniatis,T.
 Human protocadherin genes
 JOURNAL Unpublished
 TITLE
 3 (bases 1 to 5218)
 REFERENCE

AUTHORS Wu.Q. and Maniatis,T.
 TITLE Direct Submission
 JOURNAL Submitted (17-May-1999) Department of Molecular and Cellular
 Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
 02138, USA

FEATURES
 source Location/Qualifiers
 1..5218
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="3q31"
 /tissue="brain"
 1..5218
 /gene="PCDH-alpha5"
 1..2811
 /gene="PCDH-alpha5"
 /function="cell-cell adhesion"
 /codon_start=1
 /product="protocadherin alpha 5"
 /protein_id="AAB43707.1"
 /db_xref="GI:5456914"

BASE COUNT 1385 a 1207 c 1281 g 1345 t

ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 5218;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatca 25
 |||||||||||||
 Db 4115 CAAGTCACAGCAATTA 4101

RESULT 35
 AF152315 5221 bp mRNA PRI 22-JUL-1999
 LOCUS AF152315 Homo sapiens protocadherin alpha 7 (PCDH-alpha7) mRNA, complete
 DEFINITION cds.
 ACCESSION AF152315
 VERSION AF152315.1 GI:5456917
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE Wu.Q. and Maniatis,T.
 A striking organization of a large family of human neural
 cadherin-like cell adhesion genes
 JOURNAL Cell 97 (6), 779-790 (1999)
 MEDLINE 99308636
 PUBMED 10380929
 2 (bases 1 to 5221)
 Wu.Q. and Maniatis,T.
 Human protocadherin genes
 JOURNAL Unpublished
 TITLE

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5221)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA

FEATURES
source location/Qualifiers
1..5221
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue_type="brain"
1..5221
/gene="PCDH-alpha7"
1..2814
/gene="PCDH-alpha7"
/function="cell-cell adhesion"
/product="cell-cell adhesion"
/protein_id="A043709.1"
/db_xref="GI:5456918"

gene
CDS

translation="MCCPNGYDGGCHLLFTLLAAWEGSGQLHYSPYBEAKHGF
VGRADQGLLELAEVPRLEFRVAKRFGDLEVNQNLTFVNSRIDRELCGRSAEC
SILHEVIVDPLOYFHYDVEKDIINDPVPFATQKRLFLASRPIDSHFPLEGASDA
DIGENALITRYLSPNEYFLDVPISNOOVKPLGLVRLKLLDREPELHLLLTADGG
KPELGTIVOLLIVLNDNDAPVEDRLTYVRLPENVSIGTLVTHPNASDLDEGNGD
ITYSSSDVSPDIKSKFMHDLISLIPEDDAGPTVITLISVEDRDFGNCVICSILPR
TVLEVDVNDNAPOLITLSLIPEDDAGPTVITLISVEDRDFGNCVICSILPR
VPEKLVTEFKNYISLVDSADRESVATLVYVARDGSPSLMATVASYVEADVNT
NAPFAPEEYVPEKNNPPGCHLFTVYVARDGSPSLMATVASYVEADVNT
SVHSEKRYVALDPLDEHELELQFOVSARDAGVPLGSMVTLQVVLDENDAPAL
APRYGCGAVRELVPKRVAGHVAKAVADSGYNAMLSEYLQVPAAGASIPPRV
GLYGEISTRIALDETDAPRRRLVAVKDGEPSTLATATVLSVLESCQACSLVPLV
SLTAGPELTELVDVNYLLIATCAVSSLVITLLLTALRCASPSSEGCASLVKPTLV
CSSAVGSMSPSOORROVRCGEPPKTDMAFSPSLDGGSTNDPROMRYVAS
LRAGHSHSVHLEEGILRAGCGGDDQWPIVSSATPEPACGVSPPVAGVNSMTF
KYPGNPKQSGPELPPKFTIPGSPALISIRQEPNLSQIDKSDFTTFCGKEETKRRK
KKKKNKTKOEKKEKGNSTTDSQ"

BASE COUNT 1333 a 1242 c 1295 g 1351 t
ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 5221;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
Db 4118 CAAGTCACAGATT 4104

RESULT 36
AF152308 5233 bp mRNA PRI 22-JUL-1999
LOCUS AF152308/c
DEFINITION Homo sapiens protocadherin alpha 12 (PCDH-alpha12) mRNA, complete
ACCESSION CDS:
VERSION AF152308.1 GI:5456901
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 5233)
AUTHORS Wu, Q. and Maniatis, T.
TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL Cell 97 (6), 779-790 (1999)
MEDLINE 99308636
PUBMED 10380929
REFERENCE 2 (bases 1 to 5233)

AUTHORS Wu, Q. and Maniatis, T.
TITLE Human protocadherin genes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5233)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA

FEATURES
source location/Qualifiers
1..5233
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue_type="brain"
1..5233
/gene="PCDH-alpha12"
1..2826
/gene="PCDH-alpha12"
/function="cell-cell adhesion"
/product="cell-cell adhesion"
/protein_id="A043702.1"
/db_xref="GI:5456902"

gene
CDS

translation="MVIIGRGGSGRLLSLTLLAAWEGSGQLHYSPYBEAKHGF
VGRADQGLLELAEVPRLEFRVAKRFGDLEVNQNLTFVNSRIDRELCGRSAEC
SILHEVIVDPLOYFHYDVEKDIINDPVPFATQKRLFLASRPIDSHFPLEGASDA
DIGENALITRYLSPNEYFLDVPISNOOVKPLGLVRLKLLDREPELHLLLTADGG
KPELGTIVOLLIVLNDNDAPVEDRLTYVRLPENVSIGTLVTHPNASDLDEGNGD
ITYSSSDVSPDIKSKFMHDLISLIPEDDAGPTVITLISVEDRDFGNCVICSILPR
TVLEVDVNDNAPOLITLSLIPEDDAGPTVITLISVEDRDFGNCVICSILPR
VPEKLVTEFKNYISLVDSADRESVATLVYVARDGSPSLMATVASYVEADVNT
NAPFAPEEYVPEKNNPPGCHLFTVYVARDGSPSLMATVASYVEADVNT
SVHSEKRYVALDPLDEHELELQFOVSARDAGVPLGSMVTLQVVLDENDAPAL
APRYGCGAVRELVPKRVAGHVAKAVADSGYNAMLSEYLQVPAAGASIPPRV
GLYGEISTRIALDETDAPRRRLVAVKDGEPSTLATATVLSVLESCQACSLVPLV
SLTAGPELTELVDVNYLLIATCAVSSLVITLLLTALRCASPSSEGCASLVKPTLV
CSSAVGSMSPSOORROVRCGEPPKTDMAFSPSLDGGSTNDPROMRYVAS
LRAGHSHSVHLEEGILRAGCGGDDQWPIVSSATPEPACGVSPPVAGVNSMTF
KYPGNPKQSGPELPPKFTIPGSPALISIRQEPNLSQIDKSDFTTFCGKEETK
KKKKNKTKOEKKEKGNSTTDSQ"

BASE COUNT 1364 a 1213 c 1305 g 1351 t
ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 5233;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
Db 4130 CAAGTCACAGATT 4116

RESULT 37
AF152312 5251 bp mRNA PRI 22-JUL-1999
LOCUS AF152312/c
DEFINITION Homo sapiens protocadherin alpha 4 (PCDH-alpha4) mRNA, complete
ACCESSION CDS:
VERSION AF152312.1 GI:5456911
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 5251)
AUTHORS Wu, Q. and Maniatis, T.
TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL Cell 97 (6), 779-790 (1999)
MEDLINE 99308636

PUBMED 10380929
 REFERENCE 2 (bases 1 to 5251)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE Human protocadherin genes
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 5251)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA

FEATURES
 source
 Location/Qualifiers
 1..5251
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5q31"
 /tissue="brain"
 1..5251
 /gene="PCDH-alpha4"
 1..2844
 /gene="PCDH-alpha4"
 /function="cell-cell adhesion"
 /product="protocadherin alpha 4"
 /protein_id="A043706.1"
 /db_xref="GI:5456912"
 /translation="MERSGSGSGEGRLLLLLLLAWEAGNGQLHYSYEERKRGHTVGRHADDLELAEVLPRFVASKRGGLLEVNQNGILFVNSRIDRELCGRSADCSHLEIVDRPLQVFNHVEVDKINDNPPFPAOKNLISAPRIDSRRPLEGASDADEGNALTYRLSPNEYSLEKPPDELYKGLIKSLDREAPETFLVLTADGGKEELGTVOILLVLADANDAPADRTYKVRLLFNNGTLVITKLNASDLDEGNIDIVYSENDISPNWKSKEFHIDPTGQITVKGVIDPEKSEYETIVCTKQQLPSGHCRIYVEEDNDNDVPLEKSLSTIREAPLCTVAILSVSKDQWGVLTCSLTSHVPEKASTKNTYSLVDSALDRRESATLVLTARQSGSPPLMAFASVEVADVNDNAPRAQEPITYFVKEENPPGCHITFVSAMDADAQENALVSYSLVERLGRSLSSVSVHESGRVYALQPLDHELELDLOFVSARDGVPPLGSLTLQVFLDENNDAPALLARAGGTGAVSELVPSVGVHVAKRAVDAGSYNMLSYELQPGTGGRIPFRVGLYGEISTTRALDETDARHRLVLVKGHGPALATATVIVISVSGAAPKASRALVAGVPDAALVDVWVYLITACVSSILVLTLLITLALRCSALPTBACACGPKTIVPMDRYSASLRAGMHSSVHEEAGILRAGRGDPQDPVTSATPEAEVSPVQACVNSNSWTFKYGPNPKQSGELPDRFTIIPGSPALISIQEPTNSQIDKSDFTFEGKKEETKRRKKKKKKGKTKTEKKKSGNSTDNSDQ"

BASE COUNT 1360 a 1229 c 1304 g 1358 t
 ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 5251;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||||||
 Db 4148 CAAGTCACAGATT 4134

RESULT 38
 AF152306 5254 bp mRNA PRI 22-JUL-1999
 LOCUS AF152306/c
 DEFINITION Homo sapiens protocadherin alpha 10 (PCDH-alpha10) mRNA, complete
 cds
 ACCESSION AF152306
 VERSION AF152306.1 GI:5456897
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 5254)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE A striking organization of a large family of human neural cadherin-like cell adhesion genes

JOURNAL Cell 97 (6), 779-790 (1999)
 MEDLINE 99308636
 PUBMED 10380929
 REFERENCE 2 (bases 1 to 5254)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE Human protocadherin genes
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 5254)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA

FEATURES
 source
 Location/Qualifiers
 1..5254
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5q31"
 /tissue="brain"
 1..5254
 /gene="PCDH-alpha10"
 1..2847
 /gene="PCDH-alpha10"
 /function="cell-cell adhesion"
 /product="protocadherin alpha 10"
 /protein_id="A043700.1"
 /db_xref="GI:5456898"
 /translation="MVRSGCLGVOCLLLSLLLAWEAGNGQLHYSYEERKRGHTVGRHADDLELAEVLPRFVASKRGGLLEVNQNGILFVNSRIDRELCGRSADCSHLEIVDRPLQVFNHVEVDKINDNPPFPAOKNLISAPRIDSRRPLEGASDADEGNALTYRLSPNEYSLEKPPDELYKGLIKSLDREAPETFLVLTADGGKEELGTVOILLVLADANDAPADRTYKVRLLFNNGTLVITKLNASDLDEGNIDIVYSENDISPNWKSKEFHIDPTGQITVKGVIDPEKSEYETIVCTKQQLPSGHCRIYVEEDNDNDVPLEKSLSTIREAPLCTVAILSVSKDQWGVLTCSLTSHVPEKASTKNTYSLVDSALDRRESATLVLTARQSGSPPLMAFASVEVADVNDNAPRAQEPITYFVKEENPPGCHITFVSAMDADAQENALVSYSLVERLGRSLSSVSVHESGRVYALQPLDHELELDLOFVSARDGVPPLGSLTLQVFLDENNDAPALLAARAGGTGAVSELVPSVGVHVAKRAVDAGSYNMLSYELQPGTGGRIPFRVGLYGEISTTRALDETDARHRLVLVKGHGPALATATVIVISVSGAAPKASRALVAGVPDAALVDVWVYLITACVSSILVLTLLITLALRCSALPTBACACGPKTIVPMDRYSASLRAGMHSSVHEEAGILRAGRGDPQDPVTSATPEAEVSPVQACVNSNSWTFKYGPNPKQSGELPDRFTIIPGSPALISIQEPTNSQIDKSDFTFEGKKEETKRRKKKKKKGKTKTEKKKSGNSTDNSDQ"

BASE COUNT 1358 a 1210 c 1317 g 1369 t
 ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 5254;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||||||
 Db 4151 CAAGTCACAGATT 4137

RESULT 39
 AF152310 5254 bp mRNA PRI 22-JUL-1999
 LOCUS AF152310/c
 DEFINITION Homo sapiens protocadherin alpha 2 (PCDH-alpha2) mRNA, complete
 cds
 ACCESSION AF152310
 VERSION AF152310.1 GI:5456905
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 5254)
 AUTHORS Wu, Q. and Maniatis, T.

TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL Cell 97 (6), 779-790 (1999)
MEDLINE 99308636
PUBMED 10380929
REFERENCE 2 (bases 1 to 5254)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Human protocadherin genes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5254)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Direct Submission
Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA

FEATURES
source

1..5254
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue_type="brain"
1..5254
/gene="PCDH-alpha2"
1..2847
/gene="PCDH-alpha2"
/function="cell-cell adhesion"
/codon_start=1
/product="protocadherin alpha 2"
/protein_id="A043704.1"
/db_xref="GI:5456906"
/translation="MASSIRRGKGRKTRLLSLLLLAWEVSGQLRYSVPEAKHGTF
VGRADODGLELELVLPRFVASKRRHGLLEVNILONGITFVNSRLDRLELGGASAC
SIVHEVLYDRPLQVHVEVEVDINDNPIFPMVTYKIRPESRLDSRPLEGASDA
DIGNALSYRLSSSEFFLDIANDLSLSLVKSLDRETAENVLLVATDGG
KPELTGYOVLIVLVNDONEPTFAQSVYKYLEMTANGTLYVKNASDADGPNSE
IYSGDSVSTIOTKTPTIDPISGELRTKGLDYEPKSYEIOVTATDGTPTMSGHC
KISLKVINDNTPEVSTLSLPISSNASLGVIALITYSDPDGTCNGHVCSLPH
VPRVLFSTFNYSVLVDSALDRBSVAELVVTARDGSSPLMATTSSIEVADNV
NAPFAQPEYTVFVKNPPGCHITFVSADADQENALVSYSLVERGERALSSIV
SVAHSGKVAVALPDLHEVELLQFOVSARDAGVPLGSNVTLOVEVLDENDAPAL
APRAGTAGAVSELVPSVAGHVAVRAVDGYNAMLSYELQGTGSAIPPRV
GLVGTGETSTRALDEADSPRRRLIVKDHGEPALATATVLSVESGOAPASSRA
WVGAAGSEATLVNMYLITATCAVSSLVITATATLARGSVPRGGAAPCKPLV
CSSAVGWSYSGQRRORVCSGEPKTDLMAPSPSSQGPDSAPKQLSSEVYVCP
QPNPDWRTSASLRAGHSSVHLEAGILRAGPGPDQWPTVSSATPEPAGVSPV
GAGVNSWTFKYGPNPKOSGPELDPKFTIPGSPALISIRQEPNQSQDKSDFTIF
GKKEETKRRKKRKKGNKTOEKKEKGNSTTNSDO"

BASE COUNT 1375 a 1234 c 1300 g 1345 t
ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 5254;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatta 25
|||||
Db 4151 CAAGTCACAGAAATTA 4137

RESULT 40
AF152307/c 5257 bp mRNA PRT 22-JUL-1999
LOCUS Homo sapiens protocadherin alpha 11 (PCDH-alpha11) mRNA, complete
DEFINITION cds.
ACCESSION AF152307
VERSION AF152307.1 GI:5456899
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 5257)
AUTHORS Wu, Q. and Maniatis, T.
TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL Cell 97 (6), 779-790 (1999)
MEDLINE 99308636
PUBMED 10380929
REFERENCE 2 (bases 1 to 5257)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Human protocadherin genes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5257)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Direct Submission
Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA

FEATURES
source

1..5257
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue_type="brain"
1..5257
/gene="PCDH-alpha11"
1..2850
/gene="PCDH-alpha11"
/function="cell-cell adhesion"
/codon_start=1
/product="protocadherin alpha 11"
/protein_id="A043701.1"
/db_xref="GI:5456900"
/translation="MRGFORGLCTPRLOLLELLEWEVSGQLHYSVPEAKHGTF
VGRADODGLELAEVLORLFRVASKRTHGLLEVNILONGITFVNSRLDRLELGGASAC
SIVHEVLYDRPLQVHVEVEVDINDNPIFPMVTYKIRPESRLDSRPLEGASDA
DIGNALSYRLSSSEFFLDIANDLSLSLVKSLDRETAENVLLVATDGG
KPELTGYOVLIVLVNDONEPTFAQSVYKYLEMTANGTLYVKNASDADGPNSE
IYSGDSVSTIOTKTPTIDPISGELRTKGLDYEPKSYEIOVTATDGTPTMSGHC
KISLKVINDNTPEVSTLSLPISSNASLGVIALITYSDPDGTCNGHVCSLPH
VPRVLFSTFNYSVLVDSALDRBSVAELVVTARDGSSPLMATTSSIEVADNV
NAPFAQPEYTVFVKNPPGCHITFVSADADQENALVSYSLVERGERALSSIV
SVAHSGKVAVALPDLHEVELLQFOVSARDAGVPLGSNVTLOVEVLDENDAPAL
APRAGTAGAVSELVPSVAGHVAVRAVDGYNAMLSYELQGTGSAIPPRV
GLVGTGETSTRALDEADSPRRRLIVKDHGEPALATATVLSVESGOAPASSRA
WVGAAGSEATLVNMYLITATCAVSSLVITATATLARGSVPRGGAAPCKPLV
CSSAVGWSYSGQRRORVCSGEPKTDLMAPSPSSQGPDSAPKQLSSEVYVCP
QPNPDWRTSASLRAGHSSVHLEAGILRAGPGPDQWPTVSSATPEPAGVSPV
GAGVNSWTFKYGPNPKOSGPELDPKFTIPGSPALISIRQEPNQSQDKSDFTIF
GKKEETKRRKKRKKGNKTOEKKEKGNSTTNSDO"

BASE COUNT 1393 a 1224 c 1317 g 1323 t
ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 5257;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatta 25
|||||
Db 4154 CAAGTCACAGAAATTA 4140

RESULT 41
AF152305/c 5260 bp mRNA PRT 22-JUL-1999
LOCUS Homo sapiens protocadherin alpha 1 (PCDH-alpha1) mRNA, complete
DEFINITION cds.
ACCESSION AF152305
VERSION AF152305.1 GI:5456895
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
 REFERENCE
 1 (bases 1 to 5260)
 WU, Q. and Maniatis, T.
 TITLE
 A striking organization of a large family of human neural
 cadherin-like cell adhesion genes
 JOURNAL
 Cell 97 (6), 779-790 (1999)
 MEDLINE
 99308636
 PUBMED
 10380929
 REFERENCE
 2 (bases 1 to 5260)
 WU, Q. and Maniatis, T.
 TITLE
 Human protocadherin genes
 JOURNAL
 Unpublished
 3 (bases 1 to 5260)
 WU, Q. and Maniatis, T.
 TITLE
 Direct Submission
 Submitted (17-MAY-1999) Department of Molecular and Cellular
 Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
 02138, USA
 FEATURES
 source
 1..5260
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5q31"
 /tissue-type="brain"
 1..5260
 /gene="PCDH-alpha1"
 1..2853
 /gene="PCDH-alpha1"
 /function="cell-cell adhesion"
 /codon_start=1
 /product="protocadherin alpha 1"
 /protein_id="AAP43699.1"
 /db_xref="GI:5456896"
 /translation="WFSRGGGLGARDLLMLLLAAMEVSGQILHSIPEAKHGT
 VGRAVDGLLELAEVLEPRFVASKTHRLLEVNQNGILFVNSRIDEEELCGSAEC
 SHLELADRLPHEVKEVDINDNPVFEGRGOLIPESRLNRPETEGAAAD
 DIGANALLTYTLPSPDYSLDEADDELSESLMELRYIDRETPELHLLITATDGC
 KPELGTEVLEITLVNDNAPLPDQAGVARYRHILETTNGLTTLTNSDDESGNG
 VYFSPDGSISRIQKRVYDSSSGGIRLIDKUDIEKSTYEDIVKADGKSGSPSMNH
 KYLVAVLVNDNAPLAVTSLLYLPREDAPLSYVIALITVSIRDGSGNGVCTSLMH
 VPKLVSTFKNYSLVLDLSDRESLVELEVTARADGSPSLMATAVSVEADVNI
 NAPAFOPEYTVFVKEKNPPGCHIFTVSARDAQDENALVSYSLVERGERALSNV
 SVHAESGKVYALQPLDHELELLOFVSARADGSPVLSNVTLOVEVLDENDNAPLL
 APVCGTIGAVSELPRVLGAGHVAKYRAVDAGSYGNAMLSYELQAPAGARIPRY
 GLTYGELSTTRPLDEADSRRLVLYVYKHGEPALTAATVLSLVESGOAPKASRA
 SVGVAQPEALVDVNYLILAIICAVSSLLVTLTLTALRCSAPPTGCAVYPCKPTLV
 CSSALGWSNSQQRQRCVSGEGPKRTDLMASPSLPGLSPLNTSERNEQPEANLDSGN
 PRQPNFDMRYASLRAGMHSVHLEAGILRAGPGGPQOWPTVSSATPEEAGEVSP
 PVGAGVNSMTFKYGPNGPKSGELPKDFIIPGSPALISIRQEPPTNSQIDKSDFI
 TFGKKEETKRRKKKKKKGNKTQEKKEKGNSTDNDSQ"
 BASE COUNT 1365 a 1217 c 1306 g 1372 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 9; Length 5260;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caagtcacagaatta 25
 Db 4157 CAAGTCACAGATT 4143
 RESULT 42
 AF152309 5260 bp mRNA PRI 22-JUL-1999
 LOCUS AF152309/c
 DEFINITION Homo sapiens protocadherin alpha 13 (PCDH-alpha13) mRNA, complete
 CDS.
 ACCESSION AF152309.1 GI:5456903
 VERSION
 KEYWORDS

SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
 REFERENCE
 1 (bases 1 to 5260)
 WU, Q. and Maniatis, T.
 TITLE
 A striking organization of a large family of human neural
 cadherin-like cell adhesion genes
 JOURNAL
 Cell 97 (6), 779-790 (1999)
 MEDLINE
 99308636
 PUBMED
 10380929
 REFERENCE
 2 (bases 1 to 5260)
 WU, Q. and Maniatis, T.
 TITLE
 Human protocadherin genes
 JOURNAL
 Unpublished
 3 (bases 1 to 5260)
 WU, Q. and Maniatis, T.
 TITLE
 Direct Submission
 Submitted (17-MAY-1999) Department of Molecular and Cellular
 Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
 02138, USA
 FEATURES
 source
 1..5260
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5q31"
 /tissue-type="brain"
 1..5260
 /gene="PCDH-alpha13"
 1..2853
 /gene="PCDH-alpha13"
 /function="cell-cell adhesion"
 /codon_start=1
 /product="protocadherin alpha 13"
 /protein_id="AAP43703.1"
 /db_xref="GI:5456904"
 /translation="MISWGGPRPOLLMLLLAAMEVSGQILHSIPEAKHGT
 VGRADGLLELAEVLEPRFVASKTHRLLEVNQNGILFVNSRIDEEELCGSAEC
 SHLELADRLPHEVKEVDINDNPVFEGRGOLIPESRLNRPETEGAAAD
 DIGANALLTYTLPSPDYSLDEADDELSESLMELRYIDRETPELHLLITATDGC
 KPELGTEVLEITLVNDNAPLPDQAGVARYRHILETTNGLTTLTNSDDESGNG
 VYFSPDGSISRIQKRVYDSSSGGIRLIDKUDIEKSTYEDIVKADGKSGSPSMNH
 KYLVAVLVNDNAPLAVTSLLYLPREDAPLSYVIALITVSIRDGSGNGVCTSLMH
 VPKLVSTFKNYSLVLDLSDRESLVELEVTARADGSPSLMATAVSVEADVNI
 NAPAFOPEYTVFVKEKNPPGCHIFTVSARDAQDENALVSYSLVERGERALSNV
 SVHAESGKVYALQPLDHELELLOFVSARADGSPVLSNVTLOVEVLDENDNAPLL
 APVCGTIGAVSELPRVLGAGHVAKYRAVDAGSYGNAMLSYELQAPAGARIPRY
 GLTYGELSTTRPLDEADSRRLVLYVYKHGEPALTAATVLSLVESGOAPKASRA
 SVGVAQPEALVDVNYLILAIICAVSSLLVTLTLTALRCSAPPTGCAVYPCKPTLV
 CSSALGWSNSQQRQRCVSGEGPKRTDLMASPSLPGLSPLNTSERNEQPEANLDSGN
 PRQPNFDMRYASLRAGMHSVHLEAGILRAGPGGPQOWPTVSSATPEEAGEVSP
 PVGAGVNSMTFKYGPNGPKSGELPKDFIIPGSPALISIRQEPPTNSQIDKSDFI
 TFGKKEETKRRKKKKKKGNKTQEKKEKGNSTDNDSQ"
 BASE COUNT 1376 a 1234 c 1315 g 1335 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 9; Length 5260;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caagtcacagaatta 25
 Db 4157 CAAGTCACAGATT 4143
 RESULT 43
 AF152311 5260 bp mRNA PRI 22-JUL-1999
 LOCUS AF152311/c
 DEFINITION Homo sapiens protocadherin alpha 3 (PCDH-alpha3) mRNA, complete
 CDS.
 ACCESSION AF152311

```

VERSION      AF152311.1  GI:5456909
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 5260)
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE       A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
JOURNAL      Cell 97 (6), 779-790 (1999)
MEDLINE      99308636
PUBMED       10380929
REFERENCE    2 (bases 1 to 5260)
AUTHORS     Wu,Q. and Maniatis,T.
TITLE       Human protocadherin genes
JOURNAL      Unpublished
REFERENCE    3 (bases 1 to 5260)
AUTHORS     Wu,Q. and Maniatis,T.
TITLE       Direct Submission
JOURNAL      Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA

FEATURES
source       Location/Qualifiers
            1..5260
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="5"
               /map="5q31"
               /tissue="brain"
            1..5260
               /gene="PCDH-alpha3"
            1..2853
               /gene="PCDH-alpha3"
               /function="cell-cell adhesion"
               /codon_start=1
               /product="protocadherin alpha 3"
               /protein_id="AAD43705.1"
               /db_xref="GI:5456910"
               /translation="MFSMRDPDGGACILLSILLASPVSGSLHSYSEAKHGTF
               VGRADIDGIELAEIVPLPRFVASKRKHGDLLENVNLONGILFVNSRIDRELCGRSAC
               SHLEVTVPRLQVHFVEVEVDINDNAVFPMAVKNLFTSSRQSGRFSLEGASDA
               DIGTNSLITSLDSTEYFLIDVARNDEIKSLGLVAKNLNREDTPKHLITLIDGK
               KPELTGTQKLTIVLDNDNAPAFERTIKYRLLENAPNGTLVTVNATDLDEGNKD
               IASFNTDMSADILSKFHLDPYNGOISVKNIDPEKSKYELIOVEATDCKNPMSPHC
               TYLETVIDNDNPELVIOSLSPVLEDSPLSVTALISVSDRDSVNCQVCSLTPH
               VPFKLVSTFKNTYSLVLDSPLDRESATIELVYTAADGSPSLMATAVSEVADVND
               NAFVQSEYTYVFKENPPGCHITFVSARADADAPNALVSYSLVERGERALSSYV
               SVHAEQKYVALQPLDHELELLOFQVSADAGVPLGNNVLQVFLDENNAPALL
               MPVVGIGGAVSELVRSVAGHYAVKRAVDASGYNAMLSYELQPGGARIFFRV
               GLTGEISTTRALDEVDAPRIHRLVLVKHGEPSLTATATVLSLVESGQAKASSOA
               SAGATGPAALVDVNYLVAICAVSSLLVTLTLTALRCAPPTREGCGKRTLV
               CSSAVGSMYSQORQORVCSGECLEPKTIDMAFSPSPCPISPDREKQDVOVDLSAK
               PROPNDMRYASLRAGMHSVHLEAGILLAGPGGPDQWPTVSATPEPAGVESP
               PVGAGVNSNSMTFKYGGNPKQSGPELPDKFIIPGSPALISIROEPTNSQIDKSDFI
               TFGKEETKTKKKKKKKTKTOEKEKGNSTTNSDQ"

BASE COUNT   1379 a      1225 c      1293 g      1363 t
ORIGIN
Query Match 60.0%; Score 15; DB 9; Length 5260;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcaagaatla 25
    |||
Db 4157 CAAGTCACAGATTa 4143

RESULT 44
AF152314/c 5260 bp mRNA PRI 22-JUL-1999
LOCUS      Homo sapiens protocadherin alpha 6 (PCDH-alpha6) mRNA, complete
DEFINITION

```

```

cds.
ACCESSION   AF152314
VERSION     AF152314.1  GI:5456915
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE    1 (bases 1 to 5260)
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE       A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
JOURNAL      Cell 97 (6), 779-790 (1999)
MEDLINE      99308636
PUBMED       10380929
REFERENCE    2 (bases 1 to 5260)
AUTHORS     Wu,Q. and Maniatis,T.
TITLE       Human protocadherin genes
JOURNAL      Unpublished
REFERENCE    3 (bases 1 to 5260)
AUTHORS     Wu,Q. and Maniatis,T.
TITLE       Direct Submission
JOURNAL      Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA

FEATURES
source       Location/Qualifiers
            1..5260
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="5"
               /map="5q31"
               /tissue="brain"
            1..5260
               /gene="PCDH-alpha6"
            1..2853
               /gene="PCDH-alpha6"
               /function="cell-cell adhesion"
               /codon_start=1
               /product="protocadherin alpha 6"
               /protein_id="AAD43708.1"
               /db_xref="GI:5456916"
               /translation="MVFTEPDRLGKQCLLPULLLAAMKVGSGOLHSVSEAKHGTF
               VGRADIDGIELAEIVPLPRFVASKRKHGDLLENVNLONGILFVNSRIDRELCGRSAC
               SHLEVTVPRLQVHFVEVEVDINDNAVFPMAVKNLFTSSRQSGRFSLEGASDA
               DIGTNSLITSLDSTEYFLIDVARNDEIKSLGLVAKNLNREDTPKHLITLIDGK
               KPELTGTQKLTIVLDNDNAPAFERTIKYRLLENAPNGTLVTVNATDLDEGNKD
               IASFNTDMSADILSKFHLDPYNGOISVKNIDPEKSKYELIOVEATDCKNPMSPHC
               TYLETVIDNDNPELVIOSLSPVLEDSPLSVTALISVSDRDSVNCQVCSLTPH
               VPFKLVSTFKNTYSLVLDSPLDRESATIELVYTAADGSPSLMATAVSEVADVND
               NAFVQSEYTYVFKENPPGCHITFVSARADADAPNALVSYSLVERGERALSSYV
               SVHAEQKYVALQPLDHELELLOFQVSARADAGVPLGNNVLQVFLDENNAPALL
               MPVVGIGGAVSELVRSVAGHYAVKRAVDASGYNAMLSYELQPGGARIFFRV
               GLTGEISTTRALDEVDAPRIHRLVLVKHGEPSLTATATVLSLVESGQAKASSOA
               SAGATGPAALVDVNYLVAICAVSSLLVTLTLTALRCAPPTREGCGKRTLV
               CSSAVGSMYSQORQORVCSGECLEPKTIDMAFSPSPCPISPDREKQDVOVDLSAK
               PROPNDMRYASLRAGMHSVHLEAGILLAGPGGPDQWPTVSATPEPAGVESP
               PVGAGVNSNSMTFKYGGNPKQSGPELPDKFIIPGSPALISIROEPTNSQIDKSDFI
               TFGKEETKTKKKKKKKTKTOEKEKGNSTTNSDQ"

BASE COUNT   1366 a      1230 c      1315 g      1349 t
ORIGIN
Query Match 60.0%; Score 15; DB 9; Length 5260;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcaagaatla 25
    |||
Db 4157 CAAGTCACAGATTa 4143

RESULT 45
AF152316/c

```

Search completed: December 26, 2001, 12:50:03
Job time: 8725 sec

LOCUS AF152316 5260 bp mRNA P1 22-JUL-1999
DEFINITION Homo sapiens protocadherin alpha 8 (PCDH-alpha8) mRNA, complete cds.
ACCESSION AF152316
VERSION AF152316.1 GI:5456919
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 5260)
AUTHORS Wu, Q. and Maniatis, T.
TITLE A striking organization of a large family of human neural cadherin-like cell adhesion genes
JOURNAL Cell 97 (6), 779-790 (1999)
MEDLINE 99308636
PUBMED 10380929
REFERENCE 2 (bases 1 to 5260)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Human protocadherin genes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5260)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA
FEATURES
Source Location/Qualifiers
1..5260
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue_type="brain"
1..5260
/gene="PCDH-alpha8"
1..2853
/gene="PCDH-alpha8"
/function="cell-cell adhesion"
/codon_start=1
/product="protocadherin alpha 8"
/protein_id="AAD43710.1"
/db_xref="GI:5456920"
/translation="MDYHMGELGSMRLIIIIIIIAAMKVGSGLIIVSVEAKHGTFVGRJAQDLGLELAEVLRPLRVASKRHRDLLEVSLOGLIFVNSRIDRELCGRSACSIHLEVIADRPLOFVHDVEKVDNDPVRVKDQKLEVESRMPDSRPLEGASDA DVGANSVLTIRLSHDYFMDLVNSKNDENKLELVLRKSLDREDAPAHILFLATDGG KPELTGVOLLVLYLDVNDNAPTFEQSEYEVRIEFENADNGTVIKLNASDPDEGANGA IYSFNSLVEFMVTDHFSIDRNGEIVIRGNLDPEQENLYKILIDATDKGHPMAGHC TVIVRIIDKNDNPEIATLSLPIVREDAGFTVIALISVNDLDSGANGOVGCSLMPH VPEKLVSTFFKNYSLVIDSALDERRSAYELVYTARGGSPSLMATASLSVEYADYND MAPFAQPEETVEKRNPPGCHFTVSARDADQENALVSTLVERVGERLSSTYI SVHTESGRVVALDLOHELELLQFOVSARDACVPLGSNVTLOVFLDENDAPALL EPRVGTGGAASKLVPRSVGAVKAVYADVADSGYNAMLSTELQPAASSPRIPIRV GLYGEISTRTVIDEADSPRRILVLVKDGEPAATATATVLVLSVGGAPKASSRQ SAGVLGEALVDVNYLLIATCAVSSLVLTLLTALRCSALPTEGGCRACKPLV CSSAVGSMYSOOPORVSGSGEGPKTDIMAFSPCLPDLGSDVGEEDLNADHGIK PROPNPMWRKYSASLRAGMHSSVHLEBAGILRAFGCGPDQOWPTVSSATPEPPGCEVSP PVGAGVNSMSMTFTYGPNGRQSGPELPDKFTTPGSPATLSTRQPTNSQIDKSDPI TFGKRETEKRRKKRKKGNKTKQEKREKGNSTTNSDQ"
BASE COUNT 1357 a 1236 c 1310 g 1357 t
ORIGIN
Query Match 60.0%; Score 15; DB 9; Length 5260;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 11 caagtcacagaatla 25
|||||
DB 4157 CAAGTCACAGATT 4143

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:20:08 ; Search time 368.09 Seconds
(with 16 alignments)
58.228 Million cell updates/sec

Title: US-09-396-196f-1

Perfect score: 25

Sequence: 1 gacattgtcgaagtcacagaatla 25

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 33925

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

1: N_Geneseq_1101:*
2: /SID52/gcgdata/geneseq/geneseqn/NA1980.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn/NA1981.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn/NA1982.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn/NA1983.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn/NA1984.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn/NA1985.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn/NA1986.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn/NA1987.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn/NA1988.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn/NA1989.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn/NA1990.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn/NA1991.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn/NA1992.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn/NA1993.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn/NA1994.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn/NA1995.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn/NA1996.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn/NA1997.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn/NA1998.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941
2	25	100.0	1041	20	AA01303
3	25	100.0	1084	10	AA01329
4	25	100.0	1121	7	AA060496
5	25	100.0	5872	15	AA062386
6	17	68.0	17	14	AA056307
7	15	60.0	561	22	AA04235
8	15	60.0	1771	20	AA03602
9	15	60.0	1771	22	AA010264
10	15	60.0	3520	22	AA03909
11	15	60.0	3520	22	AA029361

12	C	12	15	60.0	3748	22	AAF98729	Human late stage o
13	C	13	15	60.0	6087	22	AA022683	Human cDNA encodin
14	C	14	15	60.0	6116	22	AA022447	Human cDNA encodin
15	C	15	15	60.0	6567	11	AA003324	Elmeria tenella ge
16	C	16	14	56.0	523	22	AA012853	Human cDNA clone (
17	C	17	14	56.0	1050	9	AA080927	Lividans Exported
18	C	18	14	56.0	1465	22	AA014046	Human cDNA sequenc
19	C	19	14	56.0	1677	17	AA031932	Retinoid X recepto
20	C	20	14	56.0	1787	17	AA031928	Retinoid X recepto
21	C	21	14	56.0	1890	20	AA025940	Human prostate tum
22	C	22	14	56.0	1919	22	AA060385	Human polynucleoti
23	C	23	14	56.0	2022	22	AA015029	Human cDNA sequenc
24	C	24	14	56.0	2102	22	AA015859	Human polynucleoti
25	C	25	14	56.0	2220	22	AA012786	Human nuclear horm
26	C	26	14	56.0	2233	19	AA010120	Human retinoid rec
27	C	27	14	56.0	2585	22	AA077541	Caenorhabditis bri
28	C	28	14	56.0	3083	12	AA011851	Glutamate receptor
29	C	29	14	56.0	5662	22	AA016108	Human polynucleoti
30	C	30	14	56.0	5773	21	AA076014	Human ORFX ORF1569
31	C	31	14	56.0	5876	22	AA059303	Human polynucleoti
32	C	32	14	56.0	32351	21	AA021307	Human adenosin re
33	C	33	14	56.0	32351	21	AA035185	Human low adenosin
34	C	34	14	56.0	40298	21	AA021311	Human adenosin re
35	C	35	14	56.0	40298	21	AA035189	Human low adenosin
36	C	36	14	56.0	1038602	20	AA001425	Human adenosin re
37	C	37	13	52.0	87	16	AA022576	Complete genome se
38	C	38	13	52.0	167	16	AA020233	Human gene signatu
39	C	39	13	52.0	187	21	AA026213	Human gene signatu
40	C	40	13	52.0	363	21	AA050159	Human secreted pro
41	C	41	13	52.0	378	13	AA025667	Antibody 5H7 heavy
42	C	42	13	52.0	408	21	AA074604	Sequence of the an
43	C	43	13	52.0	415	21	AA067883	Murine 5B3 antibod
44	C	44	13	52.0	476	21	AA009392	Eucahyptus grandis
45	C	45	13	52.0	566	22	AA009733	Human secreted pro
46	C	46	13	52.0	559	22	AA093531	Human cDNA clone (
47	C	47	13	52.0	603	22	AA032562	cDNA encoding SRT
48	C	48	13	52.0	632	22	AA097876	Human secreted pro
49	C	49	13	52.0	637	21	AA058530	Murine 7-transmemb
50	C	50	13	52.0	719	15	AA081957	Bovine metalloprot
51	C	51	13	52.0	735	22	AA019684	Ca
52	C	52	13	52.0	777	18	AA079355	Murine anti-human
53	C	53	13	52.0	778	20	AA032456	Staphylococcus aur
54	C	54	13	52.0	902	21	AA06756	Partial nucleotide
55	C	55	13	52.0	930	20	AA015335	Human transmembran
56	C	56	13	52.0	930	20	AA032490	cDNA encoding a le
57	C	57	13	52.0	1002	21	AA050164	Human olfactory re
58	C	58	13	52.0	1010	21	AA064778	DNA encoding 5H7 s
59	C	59	13	52.0	1010	21	AA027678	Strawberry alcohol
60	C	60	13	52.0	1278	21	AA036039	Arabidopsis thalia
61	C	61	13	52.0	1317	21	AA054840	Drosophila melanog
62	C	62	13	52.0	1359	21	AA072281	Drosophila odorant
63	C	63	13	52.0	1404	20	AA021521	Consensus phytase-
64	C	64	13	52.0	1404	21	AA032524	Consensus phytase
65	C	65	13	52.0	1404	21	AA032524	Consensus phytase
66	C	66	13	52.0	1404	21	AA032524	Consensus phytase
67	C	67	13	52.0	1404	21	AA032524	DNA encoding a mut
68	C	68	13	52.0	1404	21	AA032524	DNA encoding conse
69	C	69	13	52.0	1404	21	AA032524	Human secreted pro
70	C	70	13	52.0	1410	21	AA032524	Consensus phytase
71	C	71	13	52.0	1426	21	AA032524	DNA encoding phyt
72	C	72	13	52.0	1426	21	AA032524	DNA encoding conse
73	C	73	13	52.0	1426	21	AA032524	Human secreted pro
74	C	74	13	52.0	1490	21	AA016645	Human polynucleoti
75	C	75	13	52.0	1594	22	AA032526	Human cDNA sequenc
76	C	76	13	52.0	1600	22	AA015948	Human cDNA sequenc
77	C	77	13	52.0	1701	22	AA032528	Human fetal kidney
78	C	78	13	52.0	1713	20	AA040495	Human secreted pro
79	C	79	13	52.0	1726	20	AA040495	Human secreted pro
80	C	80	13	52.0	1887	22	AA042027	Human cDNA sequenc
81	C	81	13	52.0	1903	22	AA042027	Human cDNA sequenc
82	C	82	13	52.0	1929	21	AA033789	Human cDNA sequenc
83	C	83	13	52.0	2044	21	AA033789	Human cDNA sequenc
84	C	84	13	52.0	2058	22	AA014074	Human cDNA sequenc

```

C 85 13 52.0 2214 15 AAG5604
C 86 13 52.0 2214 20 AAG22709
C 87 13 52.0 2214 20 AAV64786
C 88 13 52.0 2214 21 AAV64786
C 89 13 52.0 2214 21 AAV64786
C 90 13 52.0 2214 21 AAV64786
C 91 13 52.0 2214 21 AAV64786
C 92 13 52.0 2214 21 AAV64786
C 93 13 52.0 2214 21 AAV64786
C 94 13 52.0 2214 21 AAV64786
C 95 13 52.0 2214 21 AAV64786
C 96 13 52.0 2214 21 AAV64786
C 97 13 52.0 2214 21 AAV64786
C 98 13 52.0 2214 21 AAV64786
C 99 13 52.0 2214 21 AAV64786
C 100 13 52.0 2214 21 AAV64786

```

ALIGNMENTS

RESULT 1

ID AAI62941 standard: DNA; 839 BP.

AC AAI62941:

DT 22-OCT-2001 (first entry)

DE Human genomic DNA SEQ ID NO 269.

XX Human: nontropic; neuroprotective; cytoskeletal; dermatological; virologic;

XX immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulvar;

XX antiparasitic; antidiabetic; antitumor; antineoplastic; cancer;

XX antiallergic; antidiabetic; antitumor; antineoplastic; cancer;

XX antiparasitic; antidiabetic; antitumor; antineoplastic; cancer;

XX neurological disease; infection; nephrotropic; gene therapy; vaccine;

XX ds.

XX Homo sapiens.

XX WO200155449-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01346.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUL-2000; 2000US-0216880.

XX 14-JUL-2000; 2000US-0218290.

XX 14-AUG-2000; 2000US-0225447.

XX 01-SEP-2000; 2000US-0229343.

XX 06-SEP-2000; 2000US-0230437.

XX 08-SEP-2000; 2000US-0231243.

XX 25-SEP-2000; 2000US-0234997.

XX 29-SEP-2000; 2000US-0236367.

XX 13-OCT-2000; 2000US-0239937.

XX 08-NOV-2000; 2000US-0246476.

XX 08-NOV-2000; 2000US-0246476.

XX 08-NOV-2000; 2000US-0246525.

XX 08-NOV-2000; 2000US-0246526.

XX 17-NOV-2000; 2000US-0249210.

XX 17-NOV-2000; 2000US-0249211.

XX 17-NOV-2000; 2000US-0249214.

XX 17-NOV-2000; 2000US-0249214.

XX 01-DEC-2000; 2000US-0250160.

XX 01-DEC-2000; 2000US-0250391.

XX 05-DEC-2000; 2000US-0251030.

XX 05-DEC-2000; 2000US-0251988.

```

PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-476225/51.
XX
XX Novel plasma membrane associated proteins useful for diagnosing,
XX treating, preventing and/or prognosing disorders related to the
XX proteins, including cancer, immune response and neuronal disorders
XX
XX Example 2: SEQ ID NO 269; 532bp + Sequence Listing: English.
XX
XX The invention relates to novel genes (AAI62752-AAI62961) and proteins
XX (AAM4347-AAM43415) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other:

```

Query Match 100.0%; Score 25; DB 22; Length 839;

Best Local Similarity 100.0%; Pred. No. 9; 9e-06;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacatgtgcgaagtcacagaatta 25

DB 53 gacatgtgcgaagtcacagaatta 77

RESULT 2

ID AAX01303 standard: DNA; 1041 BP.

XX AAX01303:

XX 12-APR-1999 (first entry)

XX E. coli biotin synthetase (BioB) coding sequence.

XX DAP aminotransferase; diaminopelargonic acid; transgenic plant;

XX biotin synthetase; biotin production; vitamin H; BioB; ss.

XX Escherichia coli.

XX US5869719-A.

XX 09-FEB-1999.

XX 30-APR-1997; 97US-0846338.

XX 30-APR-1997; 97US-0846338.

XX 08-MAR-1995; 95US-0401068.

```

PA (NOVS ) NOVARTIS FINANCE CORP.
XX
XX Palton DA:
XX
XX WPI: 1999-152902/13.
DR P-PSDB: AAW73906.
XX
XX Transgenic plants with high biotin levels - transformed with DNA
PT encoding di:amino-pelargonic acid amino-transferase or biotin
PT synthase
XX
XX Example 2: Column 37-40: 34pp: English.
XX
XX This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC be used in the transgenic plant of the invention. The transgenic plant,
CC plant cell or plant tissue is transformed with a chimeric gene encoding
CC diaminopelargonic acid (UMP) aminotransferase or biotin synthase and
CC produces more biotin than a non-transgenic plant, cell or tissue. The
CC plant is used as an improved dietary source of biotin (vitamin H) for
CC humans or animals.
XX
XX Sequence 1041 BP: 262 A; 273 C; 305 G; 201 T; 0 other;
SQ

Query Match 100.0%; Score 25; DB 20; Length 1041:
Best Local Similarity 100.0%; Pred. No. 9.9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgctgcaagtcacagaatta 25
DB 21 gacattgctgcaagtcacagaatta 45

RESULT 3
AAN91329
ID AAN91329 standard: DNA: 1084 BP.
XX
XX AAN91329:
AC
AC 15-FEB-1990 (first entry)
XX
XX E.coli Bio B gene.
DE
XX E.coli Bio B gene; biotin.
XX
XX Escherichia coli.
OS
XX
XX Key Location/Qualifiers
FH CDS 24..1064
FT /*tag=a
FT
XX
XX GB2216530-A.
XX
XX 11-OCT-1989.
PD
XX
XX 17-MAR-1989: 89GB-0006210.
PF
XX
XX 22-MAR-1988: 88GB-0006804.
PR
XX 17-MAR-1989: 89GB-0006210.
PR
XX (UKAC-) UK MIN. AGRIC. FISH.
XX
XX Pearson BM, McKee RA:
PI
XX
XX WPI: 1989-295085/41. P-PSDB: P91392
XX
XX Plasmid contg. gene(s) for: expression of biotin synthetase enzymes
PT - derived from E.coli and capable of replication and expression in other
PT microorganisms, esp. yeast.
XX
XX Table 3: page 33-4: 52pp: English.
XX
XX The gene can be used in a plasmid for expression of enzymes of the biotin
CC

```

```

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC Lactobacillus. Insertion of bio B improves biotin yields in
CC microorganisms which export biotin, or enables growth in media contg.
CC little or no biotin of organisms unable to synthesise biotin for their
CC own use.
XX
XX Sequence 1084 BP: 271 A; 286 C; 318 G; 209 T; 0 other;
SQ

Query Match 100.0%; Score 25; DB 10; Length 1084:
Best Local Similarity 100.0%; Pred. No. 9.9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgctgcaagtcacagaatta 25
DB 44 gacattgctgcaagtcacagaatta 68

RESULT 4
AAN60496
ID AAN60496 standard: DNA: 1121 BP.
XX
XX AAN60496:
AC
AC 17-OCT-1991 (first entry)
XX
XX Sequence encoding biotin synthesising enzyme.
DE
XX
XX Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
XX Key Location/Qualifiers
FH CDS 42..1082
FT /*tag= a
FT
XX
XX JP61149091-A.
XX
XX 07-JUL-1986.
PD
XX
XX 24-DEC-1984: 84JP-0272605.
PF
XX
XX 24-DEC-1984: 84JP-0272605.
PR
XX
XX (NIPS ) NIPPON SODA KK.
XX
XX WPI: 1986-216622/33.
XX
XX P-PSDB: AAP60536.
XX
XX Double stranded DNA encoding biotin synthesising enzyme.
PT comprises transformed mutant E.coli strain contg. cyclic doubled
PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
XX Disclosure: Page 534: 23pp: Japanese.
XX
XX The sequence may be expressed by a transformed E.coli host, cultured
CC in a medium containing deschiobiotin.
XX
XX Sequence 1121 BP: 290 A; 301 C; 319 G; 211 T; 0 other;
SQ

Query Match 100.0%; Score 25; DB 7; Length 1121:
Best Local Similarity 100.0%; Pred. No. 9.9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgctgcaagtcacagaatta 25
DB 62 gacattgctgcaagtcacagaatta 86

RESULT 5
AA062386
ID AA062386 standard: DNA: 5872 BP.
XX

```

AC	AA062386;	
DT	16-NOV-1994	(first entry)
XX		
DE	Biotin-biosynthesis genes contg. plasmid pB030A-15/9.	
XX		
KW	Biotin; expression: enterobacteria; vitamin H; synthesis:	
KW	plasmid; pB030A-15/9; bioc; bioc; bioc; bioc; bioa;	
KW	promoter ptac; biotin synthase; KAPA synthase;	
KW	8-amino-7-oxononanoate synthase; pimeoyl-CoA; DTB synthase;	
KW	dehbiobiotin synthase; DAPA synthase;	
KW	S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;	
KW	sedorrtioea; dermatitis; ds.	
XX		
OS	Escherichia coli DSM498.	
XX		
Key	Location/Qualifiers	
FT	promoter	1..96
FT	/*tag= a	/function= "promoter ptac"
FT	/evidence=	EXPERIMENTAL
FT	23..28	/*tag= b
FT	-35_signal	/standard_name= "promoter ptac"
FT	45..50	/*tag= c
FT	/evidence=	EXPERIMENTAL
FT	105..109	/standard_name= "promoter ptac"
FT	/*tag= d	/evidence= EXPERIMENTAL
FT	117..1157	/standard_name= "biob RBS no. 9"
FT	CDS	/*tag= e
FT	/*product=	"biotin synthase"
FT	/evidence=	EXPERIMENTAL
FT	/gene=	"biob"
FT	/number=	1
FT	1141..1146	/*tag= f
FT	CDS	/standard_name= "biob RBS"
FT	1154..2311	/*tag= g
FT	/EC_number=	2.3.1.47
FT	/product=	"KAPA synthase"
FT	/evidence=	EXPERIMENTAL
FT	/gene=	"biob"
FT	/number=	2
FT	/standard_name=	"8-amino-7-oxononanoate synthase"
FT	2284..2288	/*tag= h
FT	RBS	/standard_name= "bioc RBS"
FT	2293..3050	/*tag= i
FT	/function=	"Involved in pimeoyl-CoA synthesis"
FT	/product=	"protein"
FT	/gene=	"bioc"
FT	/number=	3
FT	3030..3033	/*tag= j
FT	/standard_name=	"biob RBS"
FT	3043..3753	/*tag= k
FT	/EC_number=	6.3.3.3
FT	/product=	"DTB synthase"
FT	/evidence=	EXPERIMENTAL
FT	/gene=	"biob15"
FT	/number=	4
FT	/standard_name=	"dehbiobiotin synthase"
FT	3712..3750	/*tag= l
FT	/note=	"biob15 substitution"
FT	3742..3746	
FT	RBS	

Query Match	100.0%; Score 25; DB 15; Length 5872;
Best local Similarity	100.0%; Pred. No. 9.9e-06;
Matches 25; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 gacattgtgcgaagtcacagaatta 25
Db	137 gacattgtgcgaagtcacagaatta 161
RESULT 6	
AA056307/c	
10	AA056307 standard; DNA; 17 BP.
XX	
CC	AA056307;


```

XX 13-APR-1994 (first entry)
XX B10B DNA primer.
XX B10A; B10B; promoter; biotin; operon; primer; ss.
XX Synthetic.
XX JP05219956-A.
XX 31-AUG-1993.
XX 14-SEP-1992: 92JP-0244792.
XX 14-SEP-1992: 92JP-0244792.
XX (SHIS ) SHISFIDO CO LTD.
XX (TAKE ) TAKEIDA CHEM IND LTD.
XX WPI: 1993-308323/39.
XX DNA sequence of biotin operon - has base sequence of E. coli
XX mutated by base pair(s) compared to wild type
XX Example 1: Fig 8; 11pp: Japanese.
XX A novel DNA sequence comprises the E.coli biotin operon (BO) in which
XX the control region of BO or the region near the b103 initiation
XX codon is mutated by at least one base pair compared to its
XX wild type. Two primers (AA056306-Q56307) are described in Example 1.
XX a microorganism belonging to Escherichia genus, transformed by
XX a recombinant plasmid carrying such DNA can be used for the prodn.
XX of biotin-active substances.
XX Sequence 17 BP: 3 A; 4 C; 4 G; 6 T; 0 other;
SQ
Query Match 68.0%; Score 17; DB 14; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 gtcgcaagtcacagaat 23
Db 17 gtcgcaagtcacagaat 1
RESULT 7
AAF94235
ID AAF94235 standard; DNA: 561 BP.
XX
AC AAF94235;
XX
DT 23-MAY-2001 (first entry)
XX
DE Primer specific for DNA encoding secretory/membrane protein SP0 ID 669.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; PCR primer; ss.
XX
OS Synthetic.
XX
PN EPI067182-A2.
XX
PD 10-JAN-2001.
XX
PE 07-JUL-2000: 2000EP-0114090.
XX
PR 08-JUL-1999: 99JP-0194179.
PR 11-JAN-2000: 2000JP-0118775.
PR 02-MAY-2000: 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.

```

```

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI WPI: 2001-093989/11.
XX
DR Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX gene therapy or as candidate target molecules in drug development -
XX
PS Claim 5; SEQ ID 669: 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
CC (ELISA). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 561 BP: 213 A; 83 C; 114 G; 149 T; 2 other;
Query Match 60.0%; Score 15; DB 22; Length 561;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 11 caagtcacagaatta 25
Db 339 caagtcacagaatta 353
RESULT 8
AAZ30602
ID AAZ30602 standard; DNA: 1771 BP.
XX
AC AAZ30602;
XX
DT 18-JAN-2000 (first entry)
XX
DE Mouse integrin alpha 4 coding sequence.
XX
KW Human; integrin; antisense; oligonucleotide; inhibition; expression;
KW very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;
KW vascular endothelial cell; vascular endothelium; migration; inflammation;
KW atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
KW metastasis; circulatory system; autoimmune disease; Grave's disease;
KW Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
XX
OS Mus sp.
XX
PN US5968826-A.
XX
PD 19-OCT-1999.
XX
PF 05-OCT-1998: 98US-0166203.
XX
PR 05-OCT-1998: 96US-0166203.

```



```

XX 07-JUL-2000; 2000EP-0114090.
PF 08-JUL-1999; 99JP-0194179.
XX 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI WPI: 2001-093989/11.
XX P-PSDB; AAB88482.
DR Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX Claim 1: SEQ ID 331; 609pp + CD ROM; English.
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
CC (ELISA)). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;
SQ

```

```

Query Match 60.0%; Score 15; DB 22; Length 3520;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 11 caagtcacagaatta 25
   |||||
DB 3182 CAAGTCACAGAATTGA 3168

```

```

RESULT 11
AAF9361/C
ID AAF29361 standard; DNA: 3520 BP.
XX AAF29361;
XX 20-APR-2001 (first entry)
XX Anyloid-beta protein agglutination regulating factor DNA SEQ ID 9.
XX Human: amyloid-beta protein; agglutination regulatory factor;
KM Alzheimer's disease; ds.
XX Homo sapiens.
OS
XX
XX WO200104299-A1.
XX
XX

```

```

PD 18-JAN-2001.
XX 06-JUL-2000; 2000WO-JP04515.
PF 08-JUL-1999; 99JP-0194179.
XX 18-OCT-1999; 99US-0159586.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Yamazaki M, Satoh S;
PI ARAKAWA H, MORITA M;
XX WPI: 2001-138347/14.
XX P-PSDB; AAB49771.
DR Polynucleotide encoding Amyloid-beta protein agglutination-controlling
PT factor, useful for inhibiting or promoting agglutination or
PT sedimentation of amyloid-beta protein and in diagnosis and screening
PT drugs for Alzheimer's disease -
XX Claim 1: Page 58-63; 72pp; Japanese.
XX This invention relates to polynucleotides AAF29357 - AAF29361 which
CC encode proteins AAB49767 - AAB49771. The proteins inhibit or promote the
CC agglutination of amyloid beta protein. The proteins and polynucleotide
CC sequences are useful in the diagnosis of Alzheimer's disease. They are
CC also useful for screening drugs which are useful for treating Alzheimer's
CC disease.
XX Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;
SQ

```

```

Query Match 60.0%; Score 15; DB 22; Length 3520;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 11 caagtcacagaatta 25
   |||||
DB 3182 CAAGTCACAGAATTGA 3168

```

```

RESULT 12
AAF98729/C
ID AAF98729 standard; DNA: 3748 BP.
XX AAF98729;
XX 02-JUL-2001 (first entry)
XX Human late stage ovarian tumour polynucleotide marker 44.
XX Human: ovarian cancer; identification; detection; characterisation;
KM tumour; kinase; marker; cytostatic; antisense gene therapy; ds.
XX Homo sapiens.
XX WO200118542-A2.
XX 15-MAR-2001.
XX 01-SEP-2000; 2000WO-US24199.
XX 03-SEP-1999; 99US-0152547.
XX 16-MAR-2000; 2000US-0190347.
XX 21-MAR-2000; 2000US-0191321.
XX 31-MAY-2000; 2000US-0208382.
XX 20-JUL-2000; 2000US-0220467.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Lee J, Thompson P, Lillie J;
XX WPI: 2001-211428/21.
XX
XX

```

XX Detection, assessment, prevention and therapy of ovarian cancer,
 PT comprises detecting changes in the expression of a variety of markers -
 XX
 PS
 PS Claim 1: Page 1196-1197; 1198pp; English.

CC The present invention describes a method for assessing whether a patient
 CC is afflicted with ovarian cancer by comparing: (1) the expression of a
 CC marker (1) (see AAF98594 to AAF98730), in a patient sample; and (2) the
 CC normal level of expression of (1) in a control non-ovarian cancer
 CC sample, where a significant difference between the level of expression
 CC in (a) and (b) is an indication that the patient is afflicted with
 CC ovarian cancer. (1) have cytostatic activities and can be used in
 CC antisense gene therapy. The method, compositions and kits from the
 CC present invention can be used for: (1) assessing and treating ovarian
 CC cancer; (2) making isolated hybridoma, which produces an antibody useful
 CC for ovarian cancer assessment; and (3) inhibiting ovarian cancer in a
 CC patient. AAF98573 to AAF98593 represent human kinase marker primers and
 CC probes which are used in the exemplification of the present invention.

SO Sequence 3748 BP; 1110 A; 825 C; 761 G; 1052 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 3748;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||
 Db 2627 CAAGTCACAGAAATTA 2613

RESULT 13
 AAS22683
 ID AAS22683 standard; cDNA; 6087 BP.

AC AAS22683;
 DT 24-OCT-2001 (first entry)

DE Human cDNA encoding a novel human protein #249.

XX
 KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnerrary; nootropic;
 KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.

OS Homo sapiens.

PN WO200155437-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US02623.

PR 25-JAN-2000; 2000US-0491404.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-451939/48.

XX P-PSDB; AA014378.

PT Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage -
 XX
 PS Claim 1: Page 676-677; 894pp; English.

CC The invention relates to polynucleotides encoding novel human

CC proteins or their active domains. The polypeptides, polynucleotides and
 CC antibodies raised against the polypeptides are used in a method of
 CC treatment of a mammal and prevention of disorders caused by the aberrant
 CC protein expression or activity. The polypeptides can be used as
 CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers, for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumour, in assays to determine biological activity, to
 CC raise antibodies/elicite an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC anti-inflammatory diseases, eczema, haemophilia, thrombosis,
 CC graft-versus-host disease, nervous system disorders, and infection.
 CC The present sequence encodes a protein of the invention.

SO Sequence 6087 BP; 1490 A; 1609 C; 1574 G; 1413 T; 1 other;

Query Match 60.0%; Score 15; DB 22; Length 6087;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 tcgcaagtcacagaa 22
 |||||
 Db 5191 tcgcaagtcacagaa 5205

RESULT 14
 AAS22447
 ID AAS22447 standard; cDNA; 6116 BP.

AC AAS22447;

DT 24-OCT-2001 (first entry)

DE Human cDNA encoding a novel human protein #13.

XX
 KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnerrary; nootropic;
 KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.

OS Homo sapiens.

PN WO200155437-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US02623.

PR 25-JAN-2000; 2000US-0491404.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-451939/48.

XX P-PSDB; AA014142.

PT Isolated polypeptides useful for treating anti-inflammatory diseases,

PT nervous system disorders, and for regenerating bone and cartilage -
 XX
 XX Claim 1; Page 190-196; 894pp; English.

CC The invention relates to polynucleotides encoding novel human
 CC proteins or their active domains. The polypeptides, polynucleotides and
 CC antibodies raised against the polypeptides are used in a method of
 CC treatment of a mammal and prevention of disorders caused by the aberrant
 CC protein expression or activity. The polypeptides can be used as
 CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers, for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumor, in assays to determine biological activity, to
 CC raise antibodies/elicit an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,
 CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence encodes a protein of the invention.

CC Sequence 6116 BP; 1491 A; 1626 C; 1582 G; 1417 T; 0 other;

QY Query Match 60.0%; Score 15; DB 22; Length 6116;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 8 tgcgaagtcacagaa 22
 ||||||||||||||||
 5220 tgcgaagtcacagaa 5234

RESULT 15
 AA003324/C
 ID AA003324 standard; DNA; 6567 BP.

XX AA003324;

DT 02-AUG-1990 (first entry)

XX Elmeria tenella genomic DNA encoding antigen GX5401FL.

XX Elmeria tenella; antigen GX5401FL; antigen GX5401; avian coccidiosis.

XX Elmeria tenella.

OS Elmeria tenella.

XX Elmeria tenella.

XX Elmeria tenella.

XX Elmeria tenella.

XX Elmeria tenella.

XX Elmeria tenella.

XX Elmeria tenella.

XX Elmeria tenella.

XX Elmeria tenella.

DR P-PSDB; AAR05222.

XX Cloned gene or fragment encoding antigenic protein -
 XX which binds with antibodies against avian coccidia, and
 XX transformed cells used in vaccine

XX Claim 10; Page 93 and Fig 14; 134pp; English.

CC An E. tenella genomic library was screened with radioactively labeled
 CC cDNA encoding the GX5401 antigen to identify clones encoding extensions
 CC of the GX5401 coding sequence. Several plaques reactive with the cDNA
 CC were identified. DNA from clone number 533 was sequenced. This analysis
 CC showed that the coding sequence in this clone overlaps with the GX5401
 CC coding sequence and extends that sequence toward the N-terminal coding
 CC region. Analysis of the new sequence together with the GX5401 antigen
 CC coding sequence reveals an open reading frame encoding an Elmeria
 CC protein of about 250,000 daltons. The protein carries several repeated
 CC peptide sequences and is rich in cysteine residues. The open reading
 CC frame encodes a potential signal sequence for protein secretion. The
 CC coding sequence for this antigenic protein is given in AA003324. Also new
 CC are an expression vector contg. the cloned gene, host cells transformed
 CC with the vector, and AP encoded by the cloned gene. The transformed cells
 CC are used in a vaccine to immunise birds against avian coccidiosis. By
 CC labelling the peptides, they can be used as a type-specific probe. The
 CC AP may also be used in an assay to detect Ab against the coccidia. The
 CC Abs are used to identify transformed cells contg. the DNA.

CC Sequence 6567 BP; 1490 A; 1741 C; 2189 G; 1147 T; 0 other;

QY Query Match 60.0%; Score 15; DB 11; Length 6567;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 gacatgtcgaagt 15
 ||||||||||||||||
 3486 GACATGTGCGCAAGT 3472

RESULT 16

ID AAH12853 standard; cDNA; 523 BP.

XX AAH12853;

DT 26-JUN-2001 (first entry)

XX Human cDNA clone (3'-primer) SEQ ID NO:9688.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

PT Primer sets for synthesizing polynucleotides, particularly the 5602

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

SQ Sequence 1465 BP; 470 A; 259 C; 270 G; 466 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1465;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatt 24
 |||
 Db 1076 CAAGTCACAGAATT 1063

RESULT 19
 AAT31932/C
 ID AAT31932 standard; DNA; 1677 BP.

AC AAT31932;

DT 27-SEP-1996 (first entry)

XX Retinoid X receptor interacting protein R1P14-2 DNA.

XX Retinoid X receptor interacting protein; RXR; R1P; R1P14-2; ss.

XX Mus sp.

XX Key Location/Qualifiers

FT CDS 1..1356
 FT 1672-1677
 FT polyA_signal /tag= a
 FT /*tag= b

XX W09621677-A1.

XX 18-JUL-1996.

XX 08-DEC-1995; 95WO-US16311.

XX 13-JAN-1995; 95US-0372652.

XX (GEHO) GEN HOSPITAL CORP.

XX Choi H, Moore D, Seol W;

XX WPI: 1996-342241/34.

XX P-PSDB: AAR99735.

XX Retinoid X receptor (RXR) interacting protein (RIP) - useful to
 PT modulate or mediate RXR function, anti-RIP antibodies can be used to
 PT determine RIP subcellular distribution patterns

XX Claim 8; Page 59-60; 90pp; English.

XX A cDNA clone (AAT31932) codes for retinoid X receptor interacting
 CC protein R1P14-2 (AAR99735), a previously undescribed orphan member

CC of the nuclear receptor superfamily which may be involved in the
 CC complex retinoid response. It was isolated from a mouse liver
 CC cDNA library using an R1P14 probe. Different R1P14-2 clones had
 CC different 5' sequences (see also AAT31935-37). Another isoform,
 CC R1P14-1, was also detected (see also AAT31928). The cDNA can be
 CC used for the recombinant prodn. of R1P14-2 in transformed host
 CC cells.

SQ Sequence 1677 BP; 486 A; 385 C; 408 G; 398 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 1677;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgtcgcaagtcac 18
 |||
 Db 630 TTGTGCAAGTCAC 617

RESULT 20

AAT31928/C

ID AAT31928 standard; DNA; 1787 BP.

AC AAT31928;

DT 27-SEP-1996 (first entry)

XX Retinoid X receptor interacting protein R1P14-1 DNA.

XX Retinoid X receptor interacting protein; RXR; R1P; R1P14-1; ss.

XX Mus sp.

XX Key Location/Qualifiers

FT CDS 12..1466
 FT 1782..1787
 FT polyA_signal /tag= a
 FT /*tag= b

XX W09621677-A1.

XX 18-JUL-1996.

XX 08-DEC-1995; 95WO-US16311.

XX 13-JAN-1995; 95US-0372652.

XX (GEHO) GEN HOSPITAL CORP.

XX Choi H, Moore D, Seol W;

XX WPI: 1996-342241/34.

XX P-PSDB: AAR99735.

XX Retinoid X receptor (RXR) interacting protein (RIP) - useful to
 PT modulate or mediate RXR function, anti-RIP antibodies can be used to
 PT determine RIP subcellular distribution patterns

XX Claim 8; Page 53-54; 90pp; English.

XX A full-length cDNA clone (AAT31928) codes for retinoid X receptor
 CC interacting protein R1P14-1 (AAR99735), a previously undescribed
 CC orphan member of the nuclear receptor superfamily which may be
 CC involved in the complex retinoid response. It was isolated from
 CC a mouse liver cDNA library using an R1P14 probe. Another isoform,
 CC R1P14-2, was also detected (see also AAT31932). The cDNA can be used
 CC for the recombinant prodn. of R1P14-1 in transformed host cells.

XX Sequence 1787 BP; 515 A; 408 C; 442 G; 422 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 1787;

Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgtgcagaatc 18
|||||
Db 740 TTGTGCGAAGTCAC 727

RESULT 21

AA252940/C

ID AA252940 standard; cDNA; 1890 BP.

XX AA252940;

DI 14-MAR-2000 (first entry)

XX Human prostate tumor cDNA library derived EST fragment #83.

XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;

XX treatment; ds.

XX Homo sapiens.

XX DE19820190-A1.

XX 04-NOV-1999.

XX 28-APR-1998; 98DE-1020190.

XX 28-APR-1998; 98DE-1020190.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX WPI: 1999-621386/54.

XX P-PSDB; AAY74061, AAY74062, AAY74063.

XX New human nucleic acid sequences from pancreatic tumors, and related

XX proteins -

XX Claim 2: Page 250; 502pp; German.

XX This invention describes novel polypeptides and their encoding nucleic

XX acids derived from human pancreatic tumor tissue which have cytostatic

XX activity. The sequences are also useful in producing pharmaceutical

XX compositions for treatment of pancreatic tumors. AA252836-253014

XX represent expressed sequence tag (EST) fragments derived from a human

XX pancreatic tumor cDNA library and which encode the proteins represented

XX in AAY73814-Y74252.

XX Sequence 1890 BP; 590 A; 358 C; 398 G; 544 T; 0 other;

XX Query Match 56.0%; Score 14; DB 20; Length 1890;

XX Best Local Similarity 100.0%; Pred. No. 27;

XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX OY 11 caagtcacagaatt 24
|||||
Db 1475 CAAATGACAGAAATT 1462

RESULT 22

AA160385

ID AA160385 standard; cDNA; 1919 BP.

XX AA160385;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4374.

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RJ.

XX WPI: 2001-442253/47.

XX P-PSDB; AAM41229.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

XX Claim 1: SEQ ID NO 4374; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and

XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,

XX immunosuppressant and cytostatic activity. The polynucleotides are useful

XX in gene therapy. A composition containing a polypeptide or polynucleotide

XX system, such as peripheral nervous injuries, peripheral neuropathy and

XX localised neuropathies and central nervous system diseases, such as

XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX utilisation of the activities such as: immune system suppression,

XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX assays for receptor activity, arthritis and inflammation, leukaemias and

XX C.N.S disorders.

XX Note: The sequence data for this patent did not form part of the printed

XX specification.

XX Sequence 1919 BP; 564 A; 394 C; 343 G; 617 T; 1 other;

XX Query Match 56.0%; Score 14; DB 22; Length 1919;

XX Best Local Similarity 100.0%; Pred. No. 27;

XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX OY 11 caagtcacagaatt 24
|||||
Db 402 caagtcacagaatt 415

RESULT 23

AAH15029/C

ID AAH15029 standard; cDNA; 2022 BP.

XX AAH15029;

DT	26-JUN-2001	(first entry)
DE	Human cDNA sequence SEQ ID NO:13001.	
XX	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.	
XX		
XX	Homo sapiens.	
XX		
XX	EP1074617-A2.	
XX		
XX	07-FEB-2001.	
XX		
XX	28-JUL-2000; 2000EP-0116126.	
XX		
XX	29-JUL-1999; 99JP-0248036.	
XX	27-AUG-1999; 99JP-0300253.	
XX	11-JAN-2000; 2000JP-0118776.	
XX	02-MAY-2000; 2000JP-0183767.	
XX	09-JUN-2000; 2000JP-0241899.	
XX		
XX	(HELI-) HELIX RES INST.	
XX		
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
XX		
XX	WPI, 2001-318749/34.	
XX		
XX	primer sets for synthesizing polynucleotides, particularly the 5602	
XX	full-length cDNAs defined in the specification, and for the detection	
XX	and/or diagnosis of the abnormality of the proteins encoded by the	
XX	full-length cDNAs -	
XX		
XX	Claim 8: SEQ ID 13001; 2537pp + CD ROM; English.	
XX		
XX	The present invention describes primer sets for synthesizing 5602	
XX	full-length cDNAs defined in the specification, where a primer set	
XX	comprises: (a) an oligo-dT primer and an oligonucleotide complementary	
XX	to the complementary strand of a polynucleotide which comprises one of	
XX	the 5602 nucleotide sequences defined in the specification, where the	
XX	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
XX	of an oligonucleotide comprising a sequence complementary to the	
XX	complementary strand of a polynucleotide which comprises a 5'-end	
XX	sequence and an oligonucleotide comprising a sequence complementary to a	
XX	polynucleotide which comprises a 3'-end sequence, where the	
XX	oligonucleotide comprises at least 15 nucleotides and the combination of	
XX	the 5'-end sequence/3'-end sequence is selected from those defined in	
XX	the specification. The primer sets can be used in antisense therapy and	
XX	in gene therapy. The primers are useful for synthesizing polynucleotides,	
XX	particularly full-length cDNAs. The primers are also useful for the	
XX	detection and/or diagnosis of the abnormality of the proteins encoded by	
XX	the full-length cDNAs. The primers allow obtaining of the full-length	
XX	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and	
XX	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to	
XX	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632	
XX	represent oligonucleotides, all of which are used in the exemplification	
XX	of the present invention.	
XX		
XX	Sequence 2022 BP; 622 A; 389 C; 418 G; 593 T; 0 other:	
XX		
XX	Query Match 56.0%; Score 14; DB 22; Length 2022;	
XX	Best Local Similarity 100.0%; Pred. No. 27;	
XX	Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
XX		
XX	0Y 11 caagtcacagaatt 24	
XX		
XX	Db 1633 CAAGTCACAGCAATT 1620	
XX		
XX	RESULT 24	
XX	AA158599/C	
XX	ID AA158599 standard; cDNA; 2102 BP.	

Query Match	Best local Similarity	Score 14:	DB 22:	Length 2102:
Matches 14:	Conservative	0:	Mismatches	0: Indels 0: Gaps 0:
QY	11 caagtcacgaatt 24	56.0%;		
DB	1564 CAAGTCACGAATT 1551			

```

RESULT 25
AAID12786/C
ID AAD12786 standard; cDNA; 2220 BP.
XX
XX AAD12786;
XX
XX 23-OCT-2001 (first entry)
XX
XX Human nuclear hormone receptor (NHR3C)-3 cDNA.
DE
XX
XX Human; nuclear hormone receptor; NHR3C-3; intracellular receptor; cancer;
KW transcriptional regulator; metabolic disorder; lysosomal storage disease;
KW Addison's disease; cystic fibrosis; endocrine disorder; atherosclerosis;
KW breast cancer; osteoporosis; Cushing's disease; reproductive disorder;
KW developmental disorder; infertility; renal tubular acidosis; cirrhosis;
KW endometriosis; hereditary neuropathy; immunological disorder; anaemia;
KW acquired immunodeficiency syndrome; AIDS; cell proliferative disorder;
KW allergy; Grave's disease; gene therapy; drug screening; cytostatic;
KW immunomodulatory; gout; diabetes; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 337..1770
FT CDS /*tag= a
FT /*product= "Human nuclear hormone receptor (NHR3C)-3"
XX
XX WC200155392-A2.
XX
XX 02-AUG-2001.
XX
XX 26-JAN-2001; 2001WO-0502741.
XX
XX 28-JAN-2000; 2000US-0178578.
XX 25-FEB-2000; 2000US-0185079.
XX 17-MAR-2000; 2000US-0189999.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Burford N, Baughn MR, Tang YF, Bandman O, Tribouley CM, Das D;
XX Policky JL, Lo TP;
XX
XX MPI; 2001-488797/53.
XX P-PSDB; AA066680.
XX
XX New human nuclear hormone receptors and polynucleotides encoding the
XX receptors, useful for diagnosing, treating and preventing metabolic,
XX endocrine, reproductive, developmental, immunological and cell
XX proliferative disorders -
XX
XX Claim 5; Page 95-96; 99pp; English.
XX
XX The present sequence is human nuclear hormone receptor (NHR3C)-3 cDNA.
XX NHR3C known as nuclear receptor or intracellular receptor, constitute
XX a protein superfamily whose members are both receptors and
XX transcriptional regulators. NHR3C is useful in the diagnosis, treatment
XX and prevention of metabolic (e.g. Addison's disease, cystic fibrosis,
XX diabetes, lysosomal storage disease), endocrine (e.g. disorders of
XX pituitary, thyroid, Cushing's disease, osteoporosis), reproductive (e.g.
XX infertility, endometriosis, breast cancer), developmental (e.g. anaemia,
XX renal tubular acidosis, hereditary neuropathy), immunological disorders
XX (e.g. acquired immunodeficiency syndrome (AIDS), allergy, Grave's
XX disease, gout) and cell proliferative disorders (e.g. atherosclerosis,
XX cirrhosis, cancer) and in the assessment of the effects of exogenous
XX compounds on the expression of nucleic acid and amino acid sequences of
XX nuclear hormone receptors. NHR3C may also be used to screen for
XX compounds that modulate the activity of NHR3C, for somatic or germ-line
XX gene therapy, to detect single nucleotide polymorphisms, as element on
XX microarray, to generate hybridisation probes and in drug screening.
XX
XX Sequence 2220 BP; 710 A; 425 C; 465 G; 620 T; 0 other;

```

```

Query Match 56.0%; Score 14; DB 22; Length 2220;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ttgtcgcaagtcac 18
DB 1041 TTGTGCGCAAGTCAC 1028

RESULT 26
AAV10120/C
ID AAV10120 standard; cDNA; 2233 BP.
XX
XX AAV10120;
XX
XX 29-MAY-1998 (first entry)
XX
XX Human retinoid receptor RRI cDNA.
DE
XX
XX Retinoid receptor; RRI; steroid receptor; agonist; antagonist; cancer;
KW adrenal deficiency; skin disorder; inflammatory disorder;
KW immune response regulator; autoimmune disease; therapeutic antibody; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 363..1781
FT CDS /*tag= a
FT /*product= RRI
FT /*note= "retinoid receptor"
XX
XX US5728548-A.
XX
XX 17-MAR-1998.
XX
XX 29-JUN-1995; 95US-0496631.
XX
XX 29-JUN-1995; 95US-0496631.
XX
XX (GENY) GENETICS INST INC.
XX
XX Bowman M;
XX
XX MPI; 1998-206567/18.
XX P-PSDB; AAW40072.
XX
XX Human retinoid receptor protein RRI - useful for, e.g. drug
XX screening, therapy and antibody production
XX
XX Claim 8; Column 11-15; 13pp; English.
XX
XX This sequence encodes a novel human steroid receptor, the retinoid
XX receptor protein or RRI. This protein can be used in screening assays
XX for steroid hormone receptor agonists and antagonists and in
XX pharmaceutical compositions for treating adrenal deficiencies, e.g.
XX Addison's disease, cancer, skin disorders, e.g. acne and psoriasis,
XX inflammatory disorders, e.g. arthritis and HIV infections. The protein
XX can also be used for regulating immune responses, e.g. as antitumour
XX agents, vaccine adjuvants, organ rejection inhibitors or agents for
XX treating autoimmune diseases. The protein can further be used to produce
XX therapeutic antibodies.
XX
XX Sequence 2233 BP; 744 A; 427 C; 463 G; 599 T; 0 other;

```

```

Query Match 56.0%; Score 14; DB 19; Length 2233;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ttgtcgcaagtcac 18
DB 1055 TTGTGCGCAAGTCAC 1042

```

```

RESULT 27
AAAF7541/C
ID AAF7541 standard; DNA: 2585 BP.
AC AAF7541;
AT 23-MAY-2001 (first entry)
XX
XX Caenorhabditis briggsae mab-21 coding sequence.
DE
XX
XX Mab-21; cell fate choice; pax-6; aniridia; Moebius syndrome;
XX chromosome 13q13; peripheral neural tissue differentiation; ds.
XX
XX Caenorhabditis briggsae.
OS
XX US6197504-B1.
XX
XX 06-MAR-2001.
XX
XX 19-JAN-1998; 98US-0008697.
XX
XX 10-APR-1996; 96US-0011607.
XX
XX 09-APR-1997; 97US-0835604.
XX
XX (CHOW/) CHOW K L.
XX
XX Chow KL;
XX
XX WPI: 2001-256362/26.
XX
XX P-PSDB: AAB74419.
XX
XX
XX Detecting expression of mab-21 gene of Caenorhabditis elegans encoding
XX novel protein required for choice of alternate cell fates in sample, by
XX contacting mRNA from sample with nucleic acid hybridizing with mab-21 -
XX
XX
XX Disclosure: column 41-44; 79pp; English.
XX
XX
XX The present invention describes a method of detecting mab-21 expression
XX in a sample involving contacting total mRNA from the sample with a probe
XX to the mab-21 gene. The sequences of the mab-21 coding sequence and
XX protein from several species are also given. The human mab-21 gene is
XX found on chromosome 13q13, a region which has been linked to Moebius
XX syndrome. The mab-21 protein shows homology to pax-6, mutations in which
XX are associated with aniridia and which is linked to peripheral neural
XX tissue differentiation.
XX
XX
XX Sequence 2585 BP; 751 A; 540 C; 559 G; 688 T; 47 other;
XX
XX
XX Query Match 56.0%; Score 14; DB 22; Length 2585;
XX Best Local Similarity 100.0%; Pred. No. 27;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 11 caagtcacagaatt 24
XX |||||||||||||||
XX Db 1058 CAAGTCACAGAATT 1045
XX
XX
XX RESULT 28
XX AAO11851/C
XX ID AAO11851 standard; DNA: 3083 BP.
XX
XX AAO11851;
XX
XX 31-JUL-1991 (first entry)
XX
XX
XX Glutamate receptor 3.
XX
XX
XX Glutamate receptor 3; probe; ligand; drug screening; ss.
XX
XX
XX Rattus rattus.
XX
XX

```

```

PH Key Location/Qualifiers
FT CDS 167..2830
FT /*tag= a
FT /product= GR3
FT sig_peptide 167..232
FT /*tag= b
FT mat_peptide 233..2830
FT /*tag= c
XX
XX
XX W09106648-A.
XX
XX
XX 16-MAY-1991.
XX
XX 25-OCT-1990; 90WO-US06153.
XX
XX 27-OCT-1989; 89US-0428116.
XX
XX (SALK ) SALK INST FOR BIOL STUD.
XX
XX
XX Heinemann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;
XX
XX WPI: 1991-164197/22.
XX
XX P-PSDB: AAR11991.
XX
XX
XX Glutamate receptors - used to screen for functional ligands and
XX identify and isolate further receptors
XX
XX
XX Disclosure: Fig 4; 109pp; English.
XX
XX
XX GluR3 cDNA was isolated from a rat forebrain cDNA using a
XX low-stringency screening protocol and a radiolabelled fragment of the
XX GluR1 cDNA as probe. The cDNA is deposited (ATCC 68133).
XX
XX The gene and protein can be used in drug screening, to
XX determine whether a substance is a functional ligand for the
XX receptor by monitoring ion channel activity.
XX
XX See also AAO11849-855.
XX
XX
XX Sequence 3083 BP; 877 A; 679 C; 735 G; 792 T; 0 other;
XX
XX
XX
XX Query Match 56.0%; Score 14; DB 12; Length 3083;
XX Best Local Similarity 100.0%; Pred. No. 27;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 3 catgtgcgaagtc 16
XX |||||||||||||||
XX Db 3034 CATGTGCGCAGTC 3021
XX
XX
XX RESULT 29
XX AAI61089/C
XX ID AAI61089 standard; cDNA: 5662 BP.
XX
XX AAI61089;
XX
XX 22-OCT-2001 (first entry)
XX
XX
XX Human polynucleotide.seq ID NO 5078.
XX
XX
XX Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX
XX Homo sapiens.
XX
XX
XX W020015312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX

```

```

XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XD, Ren F, Wang D;
PI Wang J, Zhou Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB: AAM41933.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1: SEQ ID NO 5078; 10078bp: English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SO Sequence 5662 BP; 1631 A; 1126 C; 1247 G; 1658 T; 0 other:

```

```

Query Match          56.0%; Score 14; DB 22; Length 5662;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 gcaagtcacagaat 23
   |||||
Db 4497 GCAAGTCACAGAAAT 4484

```

```

RESULT 30
AAC76014/C
ID AAC76014 standard; cDNA: 5773 BP.
XX
AC AAC76014;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
XX
XX

```

```

Human: open reading frame; ORFX: detection; cytoskeletal; hepatotropic;
vulnery; antiparkinsonian; neurotropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antinflammatory;
antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;

```

```

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN W0200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI: 2000-602362/57.
DR P-PSDB: AAB41805.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 2354-2357; 5507bp: English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytoskeletal; hepatotropic; vulnery;
CC antiparkinsonian; neurotropic; neuroprotective;
CC osteopathic; anticonvulsant; antarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy.
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SO Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other:

```

```

Query Match          56.0%; Score 14; DB 21; Length 5773;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 gcaagtcacagaat 23
   |||||
Db 4611 GCAAGTCACAGAAAT 4598

```

```

RESULT 31
AA159303/C
ID AA159303 standard; cDNA: 5876 BP.
XX
AC AA159303;
XX
DT 22-OCT-2001 (first entry)
XX

```

```

Human: open reading frame; ORFX: detection; cytoskeletal; hepatotropic;
vulnery; antiparkinsonian; neurotropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antinflammatory;
antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;

```

DE Human polynucleotide SEQ ID NO 1506.
 XX
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KM Peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KM leukemia; SS.
 XX
 XX Homo sapiens.
 OS
 PN WO200153312-A1.
 PD
 XX 26-JUL-2001.
 PD
 XX 26-DEC-2000; 2000WO-US34263.
 PF
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0523317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;
 PI
 XX WPI: 2001-442253/47.
 DR P-PSDB: AAM40147.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1: SEQ ID NO 1506; 10078bp; English.
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, Leukemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX Sequence 5876 BP; 1704 A; 1165 C; 1288 G; 1719 T; 0 other;
 SQ
 Query Match 56.0%; Score 14; DB 22; Length 5876;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 gcaagtcacagaat 23
 |||||||
 Db 4739 GCAAGTCACAGAAT 4726
 RESULT 32
 AAF21307
 ID AAF21307 standard; DNA: 32351 BP.
 XX

AC AAF21307;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Human low adenosine antisense oligonucleotide related sequence #2874.
 XX
 XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KM human; airway disorder; bronchoconstriction; lung inflammation;
 KM surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KM immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
 KM respiratory obstruction; pulmonary obstruction; impeded respiration;
 KM surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KM respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KM pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KM chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KM cancer; SS.
 KM
 OS Homo sapiens.
 OS
 PN WO200062736-A2.
 PD
 XX 26-OCT-2000.
 PD
 XX 24-MAR-2000; 2000WO-US08020.
 PF
 XX 06-APR-1999; 99US-0127958.
 PR
 XX (UYEC-) UNIV EAST CAROLINA.
 PA (NICE/) NICE J W.
 XX
 XX Nyce JW;
 PI
 XX WPI: 2000-679539/66.
 DR
 XX Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 PS
 PS Disclosure: Page 1295-1303; 1592pp; English.
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergies
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 CC
 XX Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;
 SQ

Query Match 56.0%; Score 14; DB 21; Length 32351;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 aagtcacagaatta 25
 |||
 Db 10696 aagtcacagaatta 10709

RESULT 33
 AAA35185
 ID AAA35185 standard; DNA: 32351 BP.
 XX
 AC AAA35185:
 XX
 DT 28-JUL-2000 (first entry)
 XX
 DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.
 XX
 KW Human: adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antisthmatic; cytosolic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
 OS Homo sapiens.
 XX
 PN WO200009525-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 03-AUG-1999; 99WO-US17712.
 XX
 PR 03-AUG-1998; 98US-0095212.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PI Nyce JW;
 XX
 DR WPI: 2000-205971/18.
 XX
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers -
 XX
 PS Disclosure: Page 1212-1219; 1343pp; English.
 XX
 CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antisthmatic, cytosolic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
 CC carcinomas, and cancers which may metastasise to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing the
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present

CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

SQ Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 32351;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 aagtcacagaatta 25
 |||
 Db 10696 aagtcacagaatta 10709

RESULT 34
 AAF21311
 ID AAF21311 standard; DNA: 40298 BP.
 XX
 AC AAF21311:
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Human low adenosine antisense oligonucleotide related sequence #2878.
 XX
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antisthmatic; analgesic; hypotensive; cytosolic;
 KW respiratory obstruction; pulmonary obstruction; impaired respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 OS Homo sapiens.
 XX
 PN WO200062736-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 24-MAR-2000; 2000WO-US08020.
 XX
 PR 06-APR-1999; 99US-0127958.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 XX
 PI Nyce JW;
 XX
 DR WPI: 2000-679539/66.
 XX
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX
 PS Disclosure: Page 1305-1315; 1592pp; English.
 XX
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antisthmatic, hypotensive and cytosolic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central

CC The present sequence represents the complete genome of Chlamydia
 CC trachomatis. Open reading frames (ORFs) of the genome encode
 CC polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
 CC be used to control growth of the microorganism. Chlamydia trachomatis is
 CC responsible for a large number of diseases, e.g. eye diseases such as
 CC conjunctivitis, genital diseases such as nongonococcal urethritis,
 CC epididymitis, cervicitis, salpingitis, perihepatitis, bartolinilitis;
 CC pneumopathy in breast feeding infants; and venereal
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in
 CC treating these diseases.

SO Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 56.0%; Score 14; DB 20; Length 1038602;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 atttcgcagaatca 17
 |||
 Db 736214 atttcgcagaatca 736227

RESULT 37

AAT22576
 ID AAT22576 standard; cDNA to mRNA; 87 BP.

AC AAT22576;
 DT 01-OCY-1996 (first entry)

XX Human gene signature HUMGS04192.
 DE

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.

XX Homo sapiens.
 OS

PN WO9514772-A1.
 XX

PD 01-JUN-1995.
 XX

PF 11-NOV-1994; 94WO-JP01916.
 XX

PR 12-NOV-1993; 93JP-0355504.
 XX

PA (MATS/) MATSUBARA K.
 PA (OKUBO/) OKUBO K.

PI Matsubara K, Okubo K;
 XX

DR WPI: 1995-206931/27.
 XX

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues

PS Claim 1; Page 1160; 2245pp; Japanese.
 XX

CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

SO Sequence 87 BP; 31 A; 25 C; 9 G; 21 T; 1 other;

Query Match 52.0%; Score 13; DB 16; Length 87;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagaatt 24
 |||
 Db 16 aagtcacagaatt 28

RESULT 38

AAT20233/C
 ID AAT20233 standard; cDNA to mRNA; 167 BP.

AC AAT20233;
 DT 24-JUL-1996 (first entry)

XX Human gene signature HUMGS01380.
 DE

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.

XX Homo sapiens.
 OS

PN WO9514772-A1.
 XX

PD 01-JUN-1995.
 XX

PF 11-NOV-1994; 94WO-JP01916.
 XX

PR 12-NOV-1993; 93JP-0355504.
 XX

PA (MATS/) MATSUBARA K.
 PA (OKUBO/) OKUBO K.

PI Matsubara K, Okubo K;
 XX

DR WPI: 1995-206931/27.
 XX

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues

PS Claim 1; Page 592; 2245pp; Japanese.
 XX

CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

SO Sequence 167 BP; 65 A; 23 C; 26 G; 44 T; 9 other;

Query Match 52.0%; Score 13; DB 16; Length 167;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 aagtcacagaatt 25
 118 AGTCACAGAAATT 106

RESULT 39
 AAC26213/c
 ID AAC26213 standard; cDNA; 187 BP.

AC AAC26213;
 XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 30288.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dunas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 1: SEQ ID 30288; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 187 BP; 59 A; 27 C; 30 G; 71 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 187;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagaatt 24
 163 AAGTCACAGAAATT 151

RESULT 40
 AAA50159/c
 ID AAA50159 standard; cDNA; 363 BP.

XX AAA50159;

XX 07-NOV-2000 (first entry)

DE Antibody 5H7 heavy chain variable region cDNA.

XX 5H7; monoclonal antibody; apoptosis; programmed cell death;
 KW major histocompatibility complex class I; MHC; tumour; therapy;
 KW graft-mediated immunoprotection; single chain antibody; scfv; ss.

XX Mammalia.

XX WO200047713-A2.

XX 17-AUG-2000.

PF 08-FEB-2000; 2000WO-US03234.

XX 09-FEB-1999; 99US-0119238.

XX (ARCH-) ARCH DEV CORP.

PI Woodie ES, Van Seventer JM, Kulkarni S, Kranz D, Holman P;

XX WPI: 2000-558210/51.

XX P-PSDB: AAY95776.

XX New polynucleotides encoding single chain light and heavy variable
 PT regions of a 5H7 antibody, useful in methods for conferring programmed
 PT cell death and graft-mediated immune protection of cell, tissue and
 PT organs -
 XX Claim 2: Page 36; 48pp; English.

XX The present sequence is that of cDNA coding for the heavy chain
 CC variable region (VH) (see AAY95776) of 5H7, an anti-human class I
 CC major histocompatibility complex (MHC) monoclonal antibody (MAb)
 CC which recognises a monomorphic determinant of the alpha-3 domain.
 CC MAb 5H7 has been shown to induce programmed cell death (PCD) in
 CC lymphoid tumours, peripheral blood mononuclear cells, and B6 mouse
 CC splenocytes transgenic for class I HLA-B*27. Polynucleotides
 CC encoding 5H7 VL and VH sequences were used in the construction of
 CC single chain variable immunoglobulin domains (scfv) of the 5H7
 CC antibody (see AAY95781). These can be used to confer PCD properties
 CC to cells, especially T and B lymphocyte tumour cells. Recombinant
 CC DNA molecules encoding 5H7 scfv are also used in methods for
 CC conferring graft-mediated immune protection of cell, tissue and
 CC transplanted organs.

XX Sequence 363 BP; 88 A; 101 C; 84 G; 90 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 363;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagaatt 24
 262 AAGTCACAGAAATT 250

RESULT 41
 AAQ25667/c
 ID AAQ25667 standard; cDNA; 378 BP.

XX AAQ25667;

XX 28-DEC-1992 (first entry)

KW plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bzlp; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX
 OS Eucalyptus grandis.
 PN MO200053724-A2.
 PD 14-SEP-2000.
 XX
 PF 09-MAR-2000; 2000MO-US06112.
 XX
 PR 11-MAR-1999; 99US-0266513.
 PR 18-AUG-1999; 99US-0149485.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 PI Wood M. McGrath A, Shenk M, Glenn M;
 DR WPI: 2000-579369/54.
 XX
 PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT
 PS
 XX
 CC Claim 1: Pages 513-514; 747pp; English.
 CC
 CC The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bzlp, bzlp family of G-box
 CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 CC
 SQ Sequence 415 BP; 124 A; 99 C; 88 G; 104 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 415;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 attgtcgcaagtc 16
 |||||||||||||
 DB 121 ATTGTCGCAAGTC 109

RESULT 44
 AAC09392/c
 ID AAC09392 standard; cDNA; 476 BP.
 XX
 AC AAC09392;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 13467.
 XX
 KW Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX

PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GENE-) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 DR WPI: 2000-500381/45.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PT
 PS
 XX
 CC Claim 1: SEQ ID 13467; 71pp + CD-ROM; English.
 CC
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 CC
 SQ Sequence 476 BP; 109 A; 72 C; 90 G; 200 T; 5 other;

Query Match 52.0%; Score 13; DB 21; Length 476;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagat 23
 |||||||||||||
 DB 263 CAACTCACAGAT 251

RESULT 45
 AAH09733
 ID AAH09733 standard; cDNA; 556 BP.
 XX
 AC AAH09733;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (3' primer) SEQ ID NO:6568.
 XX
 KW Human: primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

PS Claim 3; SEQ ID 6568; 2537bp + CD ROM; English.

XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 556 BP; 148 A; 144 C; 165 G; 89 T; 10 other;

Query Match

Best Local Similarity 52.0%; Score 13; DB 22; Length 556;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
 |||||
 Db 135 gcaagtcacagaa 147

Search completed: December 26, 2001, 14:16:28
 Job time: 6980 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:59:28 : Search time 143.1 Seconds
(without alignments)
39,566 Million cell updates/sec

Title: US-09-396-196f-1

Perfect score: 25

Sequence: 1 gacatgtcgaagtcacacaaatca 25

Scoring table: OLIGO.MUC

Gapop 60.0, Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size: 9

Total number of hits satisfying chosen parameters: 8663

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database: Issued Patents.NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PCITUS.COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17	68.0	17	2	US-07-944-259-2
6	15	60.0	1771	2	US-09-166-203-48
7	15	60.0	1771	4	US-09-377-309-48
8	14	56.0	1050	6	5498529-7
9	14	56.0	1677	2	US-08-372-652-14
10	14	56.0	1677	5	PCIT-US95-16311-14
11	14	56.0	1787	2	US-08-372-652-6
12	14	56.0	1787	5	PCIT-US95-16311-6
13	14	56.0	2233	1	US-08-496-631-1
14	14	56.0	2233	1	US-08-496-631-1
15	14	56.0	3083	4	US-07-718-575-5
16	14	56.0	3083	1	US-07-718-575-5
17	14	56.0	3083	2	US-08-481-206-5
18	14	56.0	3083	2	US-08-481-206-5
19	14	56.0	3083	2	US-08-481-206-5
20	14	56.0	3083	2	US-08-481-206-5
21	14	56.0	3083	2	US-08-481-206-5
22	14	56.0	3083	2	US-08-481-206-5
23	14	56.0	3083	2	US-08-481-206-5
24	14	56.0	3083	2	US-08-481-206-5
25	14	56.0	3083	2	US-08-481-206-5
26	14	56.0	3083	2	US-08-481-206-5
27	14	56.0	3083	2	US-08-481-206-5

RESULT

```

1      ? Sequence 7: Application US/08410068
2      ? Patent No. 5859335
3      ?
4      ? GENERAL INFORMATION:
5      ? APPLICANT: Patton, David
6      ? TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
7      ? NUMBER OF SEQUENCES: 14
8      ? CORRESPONDENCE ADDRESS:
9      ? ADDRESSEE: CIBA-GEIGY Corporation
10     ? STREET: 7 Skyline Drive
11     ? CITY: Hawthorne
12     ? STATE: NY
13     ? COUNTRY: USA
14     ? ZIP: 10532
15     ?
16     ? COMPUTER READABLE FORM:
17     ? MEDIUM TYPE: Floppy disk
18     ? COMPUTER: IBM PC compatible
19     ? OPERATING SYSTEM: PC-DOS/MS-DOS
20     ? SOFTWARE: Patentln Release #1.0, Version #1.30B
21     ? CURRENT APPLICATION DATA:
22     ? APPLICATION NUMBER: US/08/401,068
23     ? FILING DATE:
24     ? CLASSIFICATION: 435
25     ? PRIOR APPLICATION DATA:
26     ? APPLICATION NUMBER: US 08/351,970
27     ? FILING DATE: 08-DEC-1994
28     ? ATTORNEY/AGENT INFORMATION:
29     ? NAME: Elmer, James Scott
30     ? REGISTRATION NUMBER: 36,129
31     ? TELECOMMUNICATION INFORMATION:
32     ? TELEPHONE: 919-541-8614
33     ? TELEFAX: 919-541-8689
34     ? INFORMATION FOR SEQ ID NO: 7:
35     ? SEQUENCE CHARACTERISTICS:
36     ? LENGTH: 1041 base pairs
37     ? TYPE: nucleic acid
38     ? STRANDEDNESS: single
39     ? TOPOLOGY: linear
40     ? MOLECULE TYPE: DNA (genomic)
41     ? HYPOTHEetical: NO
42     ? FEATURE:
43     ? NAME/KEY: CDS
44     ? LOCATION: 1..1038
45     ? IDENTIFICATION METHOD: experimental
46     ? OTHER INFORMATION: /product= "Biotin synthase"
47     ?
48     ? US-08-401-068-7
49     ? /evidence= EXPERIMENTAL

```

```

Query Match      100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; P-adj No. 6.9e-06;
Matches    25; Conservative   0; Mismatches     0; Indels       0; Gaps          0;

Cy            1 gacattgtcgcgaagtcacagaatta 25
               |||||||
Db           21 GACATTGTGCGCAAGTCACAGAATTA 45

RESULT        2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719

GENERAL INFORMATION:
APPLICANT: Patcon, David
TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
```

ADDRESSSEE: No. 5869719artis Corporation
STREET: 5220 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30BA
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846,338
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38, 241
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ. ID NO. 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

```

Query Match: 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 6,9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 gacattgtcgcaagtcacagaatla 25
          |||||
Db       21 GACATTGTGCGCAGTCACAGAATTA 45

RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 117
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "Involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /EC_number= 2.6.1.62
OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biaa"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
amino:transf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "orf1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "biaa RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "orf1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

```

```

Query Match          100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6.3e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 gacattgcgaagtcacagaattta 25
Db 137 GACATTGTCGCAAGTCACAGAATTTA 161

```

```

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=1154
OTHER INFORMATION: /product="KAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="p10f"
OTHER INFORMATION: /number=2
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3043
OTHER INFORMATION: /EC_number=6.3.3.3
OTHER INFORMATION: /product="DTB synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="p10d"
OTHER INFORMATION: /number=4
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name="p10f RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name="p10d RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match
Best Local Similarity 100.0%; Score 25; DB 3; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gacattgcaagtcacagaatta 25
Db 137 GACATTGCGCAGTCACAGATT 161

RESULT 5
US-07-944-259-2/c
Sequence 2, Application US/07944259
Patent No. 5885792
```

```

GENERAL INFORMATION:
APPLICANT: Ifuku, Ohji
APPLICANT: Haze, Shuntaro
APPLICANT: Kishimoto, Jiro
APPLICANT: Nakahama, Kazuo
TITLE OF INVENTION: BIOTIN OPERON
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Wegner, Cantor, Mueller & Player
STREET: 1233 20th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-8218
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/944,259
FILING DATE: 19920914
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Cantor, Herbert I.
REGISTRATION NUMBER: 24,392
REFERENCE/DOCKET NUMBER: P-450-23557
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-0400
TELEFAX: 202-835-0605
TELEX: 440706
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-07-944-259-2

Query Match
Best Local Similarity 68.0%; Score 17; DB 2; Length 17;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 gtccgaagtcacagaatt 23
Db 17 GTCCGAAGTCACAGAATT 1

RESULT 6
US-09-166-203-48
Sequence 48, Application US/09166203A
Patent No. 5968826
GENERAL INFORMATION:
APPLICANT: Bennett, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowsett, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0323
CURRENT APPLICATION NUMBER: US/09/166,203A
CURRENT FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 60
SEQ ID NO 48
LENGTH: 1771
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1193)..(1387)
FEATURE:
```


NAME/KEY: CDS
LOCATION: (1709)..(1771)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: L20788 Genbank
DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48

Query Match 60.0%; Score 15; DB 2; Length 1771;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatt 24
|||||
DB 312 gcaagtcacagaatt 326

RESULT 7
US-09-377-309-48
Sequence 48, Application US/09377309B
Patent No. 6258790
GENERAL INFORMATION:
APPLICANT: Bennett, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowser, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0390
CURRENT APPLICATION NUMBER: US/09/377.309B
CURRENT FILING DATE: 1999-08-19
EARLIER APPLICATION NUMBER: 09/166.203
EARLIER FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 99
SEQ ID NO 48
LENGTH: 1771
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1193)..(1387)
FEATURE:
NAME/KEY: CDS
LOCATION: (1709)..(1771)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: L20788 Genbank
DATABASE ENTRY DATE: 1996-04-18
US-09-377-309-48

Query Match 60.0%; Score 15; DB 4; Length 1771;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatt 24
|||||
DB 312 gcaagtcacagaatt 326

RESULT 8
5498529-7/c
Patent No. 5498529
APPLICANT: BERKA, THOMAS R.; FORNMAID, JAMES A.; GORNIAK,
JOSEFINA G.; POSNBERG, MARTIN; STRICKLER, JAMES E.; TAYLOR, DEAN P.
TITLE OF INVENTION: PROTEIN PROTEASE INHIBITORS FROM
STREPTOMYCES
NUMBER OF SEQUENCES: 9
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162.506
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 894.167
FILING DATE: 03-JUN-1992
APPLICATION NUMBER: 346.119
FILING DATE: 16-FEB-1989

APPLICATION NUMBER: 897.245
FILING DATE: 18-AUG-1986
SEQ ID NO: 7
LENGTH: 1050
5498529-7

Query Match 56.0%; Score 14; DB 6; Length 1050;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagaatt 25
|||||
DB 251 AAGTCACAGAAATT 238

RESULT 9
US-08-372-652-14/c
Sequence 14, Application US/08372652
Patent No. 5932699
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wonil
APPLICANT: Choi, Hyeon-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
POLYPEPTIDES AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/372.652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30.162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1677 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-372-652-14

Query Match 56.0%; Score 14; DB 2; Length 1677;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgtcgcaagtcac 18
|||||
DB 630 TTGTGCGCAGATC 617

RESULT 10
PCT-US95-16311-14/c
Sequence 14, Application PC/TUS9516311
GENERAL INFORMATION:

APPLICANT: Moore, David
APPLICANT: Seol, Wongi
APPLICANT: Choi, Hwang-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16311
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO.: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1677 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-16311-14

Query Match 56.0%; Score 14; DB 5; Length 1677;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgtcgcaagtcac 18
|||||
Db 630 TTGTGCGAAGTCAC 617

RESULT 11
US-08-372-652-6/c
Sequence 6, Application US/08372652
Patent No. 5932699
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wongi
APPLICANT: Choi, Hwang-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO.: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1787 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-372-652-6

Query Match 56.0%; Score 14; DB 2; Length 1787;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgtcgcaagtcac 18
|||||
Db 740 TTGTGCGAAGTCAC 727

RESULT 12
PCT-US95-16311-6/c
Sequence 6, Application PC/TUS9516311
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wongi
APPLICANT: Choi, Hwang-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16311
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO.: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1787 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA
PCT-US95-16311-6

Query Match 56.0%; Score 14; DB 5; Length 1787;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgtcgcaagtcac 18
|||||
DB 740 TTGTGCGCAAGTCAC 727

RESULT 13
US-08-496-631-1/C
Sequence 1, Application US/08496631
Patent No. 5728548
GENERAL INFORMATION:
APPLICANT: Bowman, Michael
TITLE OF INVENTION: STEROID RECEPTOR RRI
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,631
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15248
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 363..1778
US-08-496-631-1

Query Match 56.0%; Score 14; DB 1; Length 2233;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgtcgcaagtcac 18
|||||
DB 1055 TTGTGCGCAAGTCAC 1042

RESULT 14
US-09-008-697A-7/C
Sequence 7, Application US/09008697A
Patent No. 6197504
GENERAL INFORMATION:
APPLICANT: Chow, King Lau

TITLE OF INVENTION: USES OF MAB-21
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,697A
FILING DATE: January 19, 1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chan, Albert Wai-Kit
REGISTRATION NUMBER: 36,479
REFERENCE/DOCKET NUMBER: 50752-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2585 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-008-697A-7

Query Match 56.0%; Score 14; DB 4; Length 2585;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacgaatt 24
|||||
DB 1058 CAAGTCACGAATT 1045

RESULT 15
US-07-718-575-5/C
Sequence 5, Application US/07718575
Patent No. 520257
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael NNN
APPLICANT: Bettler Ph.D., Bernhard NNN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
METHODS
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/718,575
FILING DATE: 19910813
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-07-718-575-5

Query Match 56.0%; Score 14; DB 1; Length 3083;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 catgtcgcaagtc 16
|||||
Db 3034 CATTGTCGCAGTC 3021

RESULT 16
US-08-481-206-5/c
Sequence 5, Application US/08481206
Patent No. 5739291
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael NMN
APPLICANT: Bettler Ph.D., Bernhard NMN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Preity, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,206
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,767
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-08-481-206-5

Query Match 56.0%; Score 14; DB 1; Length 3083;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 catgtcgcaagtc 16
|||||
Db 3034 CATTGTCGCAGTC 3021

RESULT 17
US-08-486-269A-5/c
Sequence 5, Application US/08486269A
Patent No. 5945509
GENERAL INFORMATION:
APPLICANT: Heinemann, Stephen F.
APPLICANT: Boulter, James R.
APPLICANT: Hollmann, Michael
APPLICANT: Bettler, Bernhard
APPLICANT: Jensen, Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
TITLE OF INVENTION: AND METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseq for Windows DEMONSTRATION Version 2.0D
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,269A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,767
FILING DATE: 04-FEB-1993
APPLICATION NUMBER: 07/718,575
FILING DATE: 21-JUN-1991
APPLICATION NUMBER: PCT/US90/06153
FILING DATE: 25-OCT-1990
APPLICATION NUMBER: 07/428,116
FILING DATE: 27-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9986
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-677-1409
TELEFAX: 619-677-1465
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: GLUR3
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 167...2830
OTHER INFORMATION:
US-08-486-269A-5

Query Match 56.0%; Score 14; DB 2; Length 3083;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 catgtcgaagtc 16
DB 3034 CATGTGCGAAGTC 3021

RESULT 18
US-08-651-572-1/c
Sequence 1, Application US/08651572
Patent No. 5789228
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,572
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Heron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-08-651-572-1

Query Match 52.0%; Score 13; DB 1; Length 1662;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
DB 974 AAGTCACAGAATT 962

RESULT 19
US-09-066-544-1/c
Sequence 1, Application US/09066544
Patent No. 6001984
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,544
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/651,572
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Heron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1744
TELEFAX: 201-994-1700
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-09-066-544-1

Query Match 52.0%; Score 13; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
DB 974 AAGTCACAGAATT 962

RESULT 20
US-08-951-086-1/c
Sequence 1, Application US/08951086
Patent No. 6074867
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:

```

MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951.086
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/651.572
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1744
TELEFAX: 201-994-1700
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-08-951-086-1

Query Match      52.0%; Score 13; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
|||||
Db 974 AAGTCACAGAATT 962

RESULT 21
US-08-765-081-4
Sequence 4, Application US/08765081
Patent No. 5798260
GENERAL INFORMATION:
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44mb storage
COMPUTER: IBM PC compatible/Pentium
OPERATING SYSTEM: MS-Windows 3.1
SOFTWARE: Word for Windows 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765.081
FILING DATE: March 26, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: CHOR-1-10286
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 4:

```

```

SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: corresponds to SEQ ID NO:1,
nucleotides 3036-5126
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli 0157: H7
STRAIN: 86-24 NALR
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2088
US-08-765-081-4

Query Match      52.0%; Score 13; DB 1; Length 2091;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 aagtcacagaatt 25
|||||
Db 692 AGTCACAGAATT 704

RESULT 22
US-09-098-082-4
Sequence 4, Application US/09098082
Patent No. 6040421
GENERAL INFORMATION:
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44mb storage
COMPUTER: IBM PC compatible/Pentium II
OPERATING SYSTEM: MS-Windows 95
SOFTWARE: Word for Windows 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098.082
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/765,081
FILING DATE: March 26, 1997
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: CHOR-1-12402
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Corresponds to SEQ ID NO:1,

```

DESCRIPTION: nucleotides 3036-5126
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli 0157: H7
STRAIN: 86-24 NALR
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2088
US-09-098-082-4

Query Match 52.0%; Score 13; DB 3; Length 2091;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 aqtcacagaatta 25
|||||
DB 692 ACTCAGACAGATTA 704

RESULT 23
PCT-US95-06994-4
Sequence 4, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILIP I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI 0157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINNESS PLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-10591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: corresponds to SEQ ID NO:1,
DESCRIPTION: nucleotides 3036-5126
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli 0157: H7

STRAIN: 86-24 NALR
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2088
PCT-US95-06994-4

Query Match 52.0%; Score 13; DB 5; Length 2091;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 aqtcacagaatta 25
|||||
DB 692 ACTCAGACAGATTA 704

RESULT 24
US-08-484-993B-1/c
Sequence 1, Application US/08484993B
Patent No. 5837497
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunoreception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,993B
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide

LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-993B-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gccagtcacagaa 22
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 25
US-08-484-158B-1/c
Sequence 1, Application US/08484158B
Patent No. 5976545
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
TITLE OF INVENTION: Pharmaceutical Compositions for
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESS: Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,158B
FILING DATE: 07-JUNE-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 09-NOV-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-92
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 32794
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO

ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-158B-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gccagtcacagaa 22
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 26
US-08-484-596A-1/c
Sequence 1, Application US/08484596A
Patent No. 5961228
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
TITLE OF INVENTION: Materials and Methods for Immunoreception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,596A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 11-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-596A-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
DB 1898 GCAAGTCACAGAA 1886

RESULT 27
US-08-480-150A-1/c
Sequence 1, Application US/08480150A
Patent No. 5989550

GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
TITLE OF INVENTION: Materials and Methods for Immuncontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480.150A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/149,223
FILING DATE: 09-NOV-1993
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31445
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-480-150A-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
DB 1898 GCAAGTCACAGAA 1886

RESULT 28
US-08-458-731-1/c
Sequence 1, Application US/08458731
Patent No. 6001599

GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
TITLE OF INVENTION: Materials and Methods for Immuncontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458.731
FILING DATE: 09-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653

TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploid
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-458-731-1

Query Match 52.0%; Score 13; DB 3; Length 2214;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
DB 1898 GCAAGTCACAGAA 1886

RESULT 29
US-08-149-223A-1/c
Sequence 1, Application US/08149223A
Patent No. 6027727
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,223A
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107

REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploid
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-149-223A-1

Query Match 52.0%; Score 13; DB 3; Length 2214;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
DB 1898 GCAAGTCACAGAA 1886

RESULT 30
US-08-484-993B-9/c
Sequence 9, Application US/08484993B
Patent No. 5837497
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,993B
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992

ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-993B-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
DB 2098 GCAAGTCACAGAA 2086

RESULT 31
US-08-484-158B-9/c
Sequence 9, Application US/08484158B
Patent No. 5976545
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Pharmaceutical Compositions for
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,158B
FILING DATE: 07-JUNE-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 09-NOV-93
PRIOR APPLICATION DATA: 08/012,990
FILING DATE: 29-JAN-93
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-92
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 32794
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-158B-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
DB 2098 GCAAGTCACAGAA 2086

RESULT 32
US-08-484-596A-9/c
Sequence 9, Application US/08484596A
Patent No. 5981228
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunoreception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,596A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 11-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:

NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-596A-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcaagtcacagaa 22
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 33
US-08-480-150A-9/C
Sequence 9, Application US/08480150A
Patent No. 5989550
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,150A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/149,223
FILING DATE: 09-NOV-1993
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.

REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-480-150A-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcaagtcacagaa 22
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 34
US-08-458-731-9/C
Sequence 9, Application US/08458731
Patent No. 6001599
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,731
FILING DATE: 09-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-458-731-9

Query Match 52.0%; Score 13; DB 3; Length 2381;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||

DB 2098 GCAAGTCACAGAA 2086

RESULT 35
US-08-149-223A-9/c
Sequence 9, Application US/08149223A
Patent No. 6027727
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunoccontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstleing, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,223A
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-149-223A-9

Query Match 52.0%; Score 13; DB 3; Length 2381;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||

DB 2098 GCAAGTCACAGAA 2086

RESULT 36
US-08-743-637B-35/c
Sequence 35, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: OUELLETTE & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586, 90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5591
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 4500 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM SOURCE:
ORGANISM: Streptococcus pneumoniae
US-08-743-637B-35

Query Match 52.0%; Score 13; DB 2; Length 4500;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagatt 24
|||||
Db 1911 AAGTCACAGATT 1899

RESULT 37
US-08-526-840B-35/C
Sequence 35, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 4500 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
US-08-526-840B-35

Query Match 52.0%; Score 13; DB 3; Length 4500;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagatt 24
|||||
Db 1911 AAGTCACAGATT 1899

RESULT 38
US-08-765-081-1
Sequence 1, Application US/08765081
Patent No. 5798260
GENERAL INFORMATION:
APPLICANT: TART, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44mb storage
COMPUTER: IBM PC compatible/Pentium
OPERATING SYSTEM: MS-Windows 3.1
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,081
FILING DATE: March 26, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: CHOR-1-10286
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli 0157:H7
STRAIN: 86-24 NALR
IMMEDIATE SOURCE:
CLONE: PEAR
US-08-765-081-1

Query Match 52.0%; Score 13; DB 1; Length 8041;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 aagtcacagatt 25
|||||
Db 3727 AAGTCACAGATT 3739

RESULT 39
US-09-098-082-1
Sequence 1, Application US/09098082
Patent No. 6040421
GENERAL INFORMATION:
APPLICANT: TART, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage
COMPUTER: IBM PC compatible/Pentium II
OPERATING SYSTEM: MS-Windows 95
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,082
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/765,081
FILING DATE: March 26, 1997
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: CHOR-1-12402
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
IMMEDIATE SOURCE:
CLONE: PEAR
US-09-098-082-1

Query Match 52.0%; Score 13; DB 3; Length 8041;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 agtcacagaatta 25
|||||

Db 3727 AGTCACAGAATTA 3739

RESULT 40
PCT-US95-06994-1
Sequence 1, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILLIP I
APPLICANT: BLICE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESS, PLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE

STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
IMMEDIATE SOURCE:
CLONE: PEAR
PCT-US95-06994-1

Query Match 52.0%; Score 13; DB 5; Length 8041;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 agtcacagaatta 25
|||||

Db 3727 AGTCACAGAATTA 3739

RESULT 41
US-09-009-913-1/c
Sequence 1, Application US/09009913
Patent No. 6087485
GENERAL INFORMATION:
APPLICANT: AXYS Pharmaceuticals, Inc.
TITLE OF INVENTION: Asthma Related Genes
NUMBER OF SEQUENCES: 339
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Ave, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,913
FILING DATE: 21-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 72928 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match
Best Local Similarity 52.0%; Score 13; DB 3; Length 72928;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
|||||
DB 26365 AAGTCACAGAATT 26353

RESULT 42
US-09-078-294-4/c
Sequence 4, Application US/09078294
Patent No. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sart, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT APPLICATION NUMBER: US/09/078,294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 80246
TYPE: DNA
ORGANISM: Nucleotide sequence of NC-contlig
US-09-078-294-4

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 80246;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
|||||
DB 5797 AAGTCACAGAATT 5785

RESULT 43
US-09-078-294-3/c
Sequence 3, Application US/09078294
Patent No. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sart, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT APPLICATION NUMBER: US/09/078,294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 3
LENGTH: 80595
TYPE: DNA
ORGANISM: Nucleotide sequence of HC-contlig
US-09-078-294-3

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 80595;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
|||||
DB 6081 AAGTCACAGAATT 6069

RESULT 44
US-09-014-065-15/c
Sequence 15, Application US/09014065
Patent No. 6033854
GENERAL INFORMATION:
APPLICANT: Kurnit, David M.
APPLICANT: Chiang, Pei-Wen
APPLICANT: Wang, Chang-Ning J.
TITLE OF INVENTION: METHOD FOR DETERMINING THE COPY NUMBER OF A NUCLEIC ACID SEQUE
FILE REFERENCE: 06498/004001
CURRENT APPLICATION NUMBER: US/09/014,065
CURRENT FILING DATE: 1998-01-27
EARLIER APPLICATION NUMBER: US 08/434,474
EARLIER FILING DATE: 1995-05-04
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 24
TYPE: DNA
ORGANISM: Homo sapiens
US-09-014-065-15

Query Match
Best Local Similarity 48.0%; Score 12; DB 3; Length 24;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacaga 21
|||||
DB 12 GCAAGTCACAGA 1

RESULT 45
PCT-US93-03077-7
Sequence 7, Application PC/TUS9303077
GENERAL INFORMATION:
APPLICANT: Board of Regents, The University of Texas System
APPLICANT: Gaynor, Richard B.
APPLICANT: Wu, Foon Kin
TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
REGULATING GENE EXPRESSION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03077
FILING DATE: 19930331


```

: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/862,025
: FILING DATE: April 2, 1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Kammerer, Patricia A.
: REGISTRATION NUMBER: 29,775
: REFERENCE/DOCKET NUMBER: UTPD270PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 713-787-1340
: TELEFAX: 713-749-2679
:
: TELEX:
: INFORMATION FOR SEQ. ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 255 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
:
: PCT-US93-03077-7

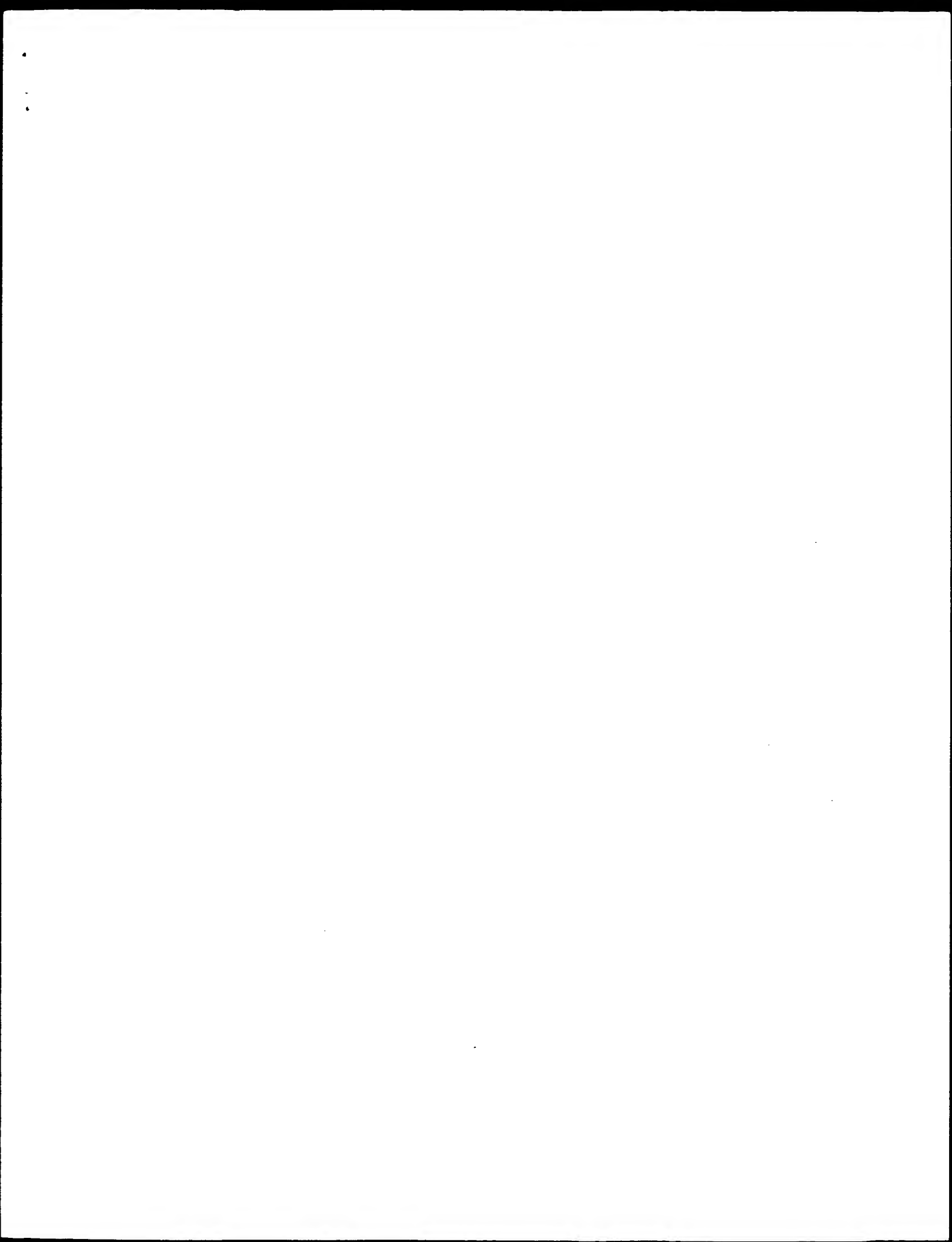
```

```

Query Match      48.08; Score 12; DB 5; Length 255;
Best Local Similarity 100.08; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 13 agtcacagaatt 24
    |||||||
Db 206 AGTCACAGAATT 217

```

Search completed: December 26, 2001, 12:52:29
Job time: 6781 sec



Thu Dec 27 14:47:24 2001

us-09-396-196f-1.olig.rst

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:12:33 ; Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196f-1

Perfect score: 25

Sequence: 1 gacattgcgcagtcacacacatla 25

Scoring table: OLIGO-MWC

Gapop 60.0 , Gapext 60.0

Word size : 9

Total number of hits satisfying chosen parameters: 531357

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthm:*
3: em_estln:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_estl:*
11: gb_estl2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hnm:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	68.0	357	11	BE800706 sq97c06.y
2	17	68.0	388	10	AW781150 s189q06.y
3	17	68.0	751	10	BE658717 CM700007A
4	17	68.0	1318	11	BE658717 602792574
5	16	64.0	299	10	BB450863 BB450863
6	16	64.0	418	11	BB450863 BB450863
7	15	60.0	172	11	BB450863 BB450863
8	15	60.0	214	10	BB450863 BB450863
9	15	60.0	233	10	BB450863 BB450863
10	15	60.0	253	10	BB450863 BB450863
11	15	60.0	256	10	BB450863 BB450863
12	15	60.0	261	11	BB450863 BB450863

C	13	60.0	263	10	BB141316	BB141316
C	14	60.0	263	11	BF745516	BF745516 CM2-BT082
C	15	60.0	282	11	BF800965	BF800965 OVA-CT010
C	16	60.0	286	10	AA344205	AA344205 EST50086
C	17	60.0	292	13	B53227	B53227 CIT-HSP-200
C	18	60.0	317	11	BF743951	BF743951 CM2-BT082
C	19	60.0	324	13	BH121038	BH121038 RPT-24-2
C	20	60.0	327	11	T30901	T30901 EST24536 Hu
C	21	60.0	335	11	BF743964	BF743964 CM2-BT082
C	22	60.0	346	11	BF743954	BF743954 CM2-BT082
C	23	60.0	349	11	BF743954	BF743954 CM2-BT082
C	24	60.0	353	10	AA648449	AA648449 ESTED25
C	25	60.0	370	11	T07051	T07051 EST04940 Fe
C	26	60.0	402	10	AI694429	AI694429 w083106.x
C	27	60.0	404	10	AI203847	AI203847 q175h03.x
C	28	60.0	405	11	BF726110	BF726110 b01c01.x
C	29	60.0	406	11	BF726110	BF726110 b01c01.x
C	30	60.0	416	10	AI942464	AI942464 w080c05.x
C	31	60.0	420	10	AI172188	AI172188 z029f06.s
C	32	60.0	427	13	AO142435	AO142435 HS-3064_A
C	33	60.0	429	10	AM870882	AM870882 r051h09.y
C	34	60.0	454	13	AO086771	AO086771 HS-2175_B
C	35	60.0	455	10	AA973493	AA973493 o045e05.s
C	36	60.0	456	10	AI756736	AI756736 ESTE2a24
C	37	60.0	457	10	AI161386	AI161386 qy47f01.x
C	38	60.0	464	11	BF562271	BF562271 ESTE2e05
C	39	60.0	471	10	AI635259	AI635259 t279h07.x
C	40	60.0	485	11	BG020076	BG020076 dc47c07.x
C	41	60.0	487	11	BG516037	BG516037 ESTED48
C	42	60.0	489	11	BF775935	BF775935 286217 MA
C	43	60.0	498	13	A2398330	A2398330 1M0163M15
C	44	60.0	502	10	AA846453	AA846453 a156b03.s
C	45	60.0	503	10	AI672463	AI672463 wa03e05.x
C	46	60.0	509	10	AI634722	AI634722 t209e04.x
C	47	60.0	512	10	AA418528	AA418528 x132c07.x
C	48	60.0	512	13	AO801291	AO801291 HS-2088_A
C	49	60.0	520	13	AO602419	AO602419 HS-2128_A
C	50	60.0	531	10	BE043069	BE043069 ho32e02.x
C	51	60.0	531	13	AO806020	AO806020 HS-3237_A
C	52	60.0	552	10	AI375298	AI375298 MCB12F10
C	53	60.0	554	10	AI268404	AI268404 qm05e10.x
C	54	60.0	562	13	AO483592	AO483592 RPT-11-2
C	55	60.0	563	13	AO430099	AO430099 HS-5074_A
C	56	60.0	587	13	A2554645	A2554645 RPT-23-1
C	57	60.0	601	10	AM828925	AM828925 ra69g07.y
C	58	60.0	617	10	AM316985	AM316985 xx10c10.x
C	59	60.0	626	10	AV657973	AV657973 AV657973
C	60	60.0	652	13	AO541562	AO541562 RPT-11-3
C	61	60.0	657	11	BF726111	BF726111 b01c01.y
C	62	60.0	658	13	A2816912	A2816912 2M0085C19
C	63	60.0	665	13	A2459164	A2459164 1M0263K18
C	64	60.0	691	13	A2963904	A2963904 2M0233D23
C	65	60.0	691	10	AA626775	AA626775 ad09e07.s
C	66	60.0	1146	11	BC174876	BC174876 602336157
C	67	60.0	1239	10	BE735021	BE735021 601569755
C	68	60.0	61	10	AI174176	AI174176 uh98f03.r
C	69	60.0	114	11	BF987752	BF987752 CM0-6N016
C	70	60.0	116	11	BF991166	BF991166 CM0-6N016
C	71	60.0	132	13	AZ073992	AZ073992 RPT-23-4
C	72	60.0	164	10	BE090319	BE090319 RC3-BT071
C	73	60.0	182	11	BI416828	BI416828 hsp002xg
C	74	60.0	182	13	AZ074961	AZ074961 RPT-23-4
C	75	60.0	190	11	F07940	F07940 HSC2M061.n
C	76	60.0	192	13	A2637253	A2637253 1M0496102
C	77	60.0	193	13	AO542941	AO542941 RPT-11-3
C	78	60.0	197	10	AA957877	AA957877 UT-R-E1-f
C	79	60.0	201	10	AA957877	AA957877 AV530463
C	80	60.0	204	10	BE684330	BE684330 184699 MA
C	81	60.0	205	10	AM579956	AM579956 PM2-HT035
C	82	60.0	205	11	BF926840	BF926840 ILS-HT022
C	83	60.0	226	10	AV345479	AV345479 AV345479
C	84	60.0	255	11	BF353051	BF353051 IL3-HT061
C	85	60.0	255	13	A2559921	A2559921 RPT-23-2

86	14	56.0	259	10	AA766388	oa36c12.s
87	14	56.0	264	10	AA329990	EST33660
88	14	56.0	267	13	B41289	HS-1053-B1-
89	14	56.0	268	11	L44566	HUMEST6D8.H
90	14	56.0	269	10	AA984664	an87402.s
91	14	56.0	270	10	BE181705	CV1-H706.3
92	14	56.0	270	13	AQ532982	RPCI-11-3
93	14	56.0	276	13	AQ414837	RPCI-11-1
94	14	56.0	276	13	AQ414837	RPCI-11-1
95	14	56.0	282	11	BG049539	EM1.5.D06
96	14	56.0	287	13	AQ415921	RPCI-11-1
97	14	56.0	289	13	AQ070825	HS-2256.A
98	14	56.0	290	11	BF804635	KC2-C1008
99	14	56.0	290	13	AO634826	RPCI-11.4
100	14	56.0	301	13	AQ352850	CITR1-EL-

ALIGNMENTS

RESULT 1
 BE800706 357 bp mRNA EST 20-SEP-2000
 LOCUS s997c06.y1 Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-c1049-1091 5' similar to TR:082150 082150 CHLOROPLAST FTSH
 PROTEASE : mRNA sequence.
 BE800706
 BE800706.1 GI:10231818

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 soybean.
 Glycine max
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
 Rosidae: eustosids I; Fabales: Fabaceae: Papilionoideae: Phaseoleae:
 Glycine.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 1 (bases 1 to 357)
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Putative full length read
 vector to vector length is 364
 High quality sequence stop: 340.

FEATURES

source
 1..357
 location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1049-1091"
 /clone_lib="Gm-c1049"
 /tissue_type="whole seedlings of greenhouse grown plants"
 /dev_stage="3 week old"
 /lab_host="DH10B"
 /note="Vector: p Bluescript II SK+; Site_1: EcoRI; Site_2:
 XhoI; The Clark NTL was constructed and seed was provided
 by Dr. J. Specht, University of Nebraska (Shoemaker and
 Specht, 1995). The cDNA library was constructed from mRNA
 isolated from whole seedlings of 3 week old greenhouse

grown plants. Complementary DNA was synthesized from mRNA
 using a primer consisting of a poly(dT) sequence with a
 XhoI restriction site and a 3' anchor. EcoRI adapters were
 ligated to the blunt-ended cDNA fragments followed by XhoI
 digestion. The cDNA fragments were directionally cloned
 into the EcoRI-XhoI restriction site of the pBluescript
 vector. The ligated cDNA fragments were transformed into
 DH10B host cells (GibcoBRL). The library was constructed
 in cooperation with Dr. Paul Keim's Laboratory at Northern
 Arizona University."

BASE COUNT
 ORIGIN
 169 a 53 c 47 g 88 t
 Query Match
 Best Local Similarity 100.0%; Pred. No. 9.1;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 ccgaagtcacagaatta 25
 Db 138 CGCAAGTCACAGATT 154

RESULT 2
 AW781150 388 bp mRNA EST 12-MAY-2000
 LOCUS s189g06.y1 Gm-c1037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-c1037-1235 5' similar to TR:082150 082150 CHLOROPLAST FTSH
 PROTEASE : mRNA sequence.
 AW781150
 AW781150.1 GI:7795753

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 soybean.
 Glycine max
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
 Rosidae: eustosids I; Fabales: Fabaceae: Papilionoideae: Phaseoleae:
 Glycine.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 1 (bases 1 to 388)
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 375.

FEATURES

source
 1..388
 location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1037-1235"
 /clone_lib="Gm-c1037"
 /tissue_type="fully expanded leaves of greenhouse grown
 plants"
 /dev_stage="2 week old"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
 cDNA library was constructed from mRNA isolated from fully
 expanded leaves of greenhouse grown plants that were 2

weeks old. The library was prepared using the Life Technologies superscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electro-Max DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. email: l-vodkin@uiuc.edu"

BASE COUNT 123 a 67 c 92 g 106 t

ORIGIN

Query Match 68.0%; Score 17; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 cgcagtcacagaatta 25
|||||

Db 334 CGCACTCAGACAAATTA 350

RESULT 3
BE658717 751 bp mRNA EST 24-MAY-2001
LOCUS GM700007A10F9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3'
DEFINITION mRNA sequence.

ACCESSION BE658717
VERSION BE658717.1 GI:9984609
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

REFERENCE

1 (bases 1 to 751)
Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V., Erpelting, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H. A functional genomics program for soybean (NSF 9872565)

UNPUBLISHED (1999)
Other ESTs: AM156684 corresponding to Gm-cl015-2632 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)

Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582

Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTT(A/C/G)-3'

FEATURES

SOURCE

1. 751
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1070-2561"

/note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or

a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the Laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, <http://www.cbc.umn.edu/ResearchProjects/soybean/index.html>. Rerecking was performed by Genome Systems, St. Louis, <http://www.genomesystems.com>, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, <http://www.life.uiuc.edu/biotech/keck.html>. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'.

BASE COUNT 193 a 171 c 131 g 226 t 30 others

ORIGIN

Query Match 68.0%; Score 17; DB 10; Length 751;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 cgcagtcacagaatta 25
|||||

Db 112 CGCACTCAGACAAATTA 96

RESULT 4
BG871816 1318 bp mRNA EST 29-MAY-2001
LOCUS 602792574F1 NCI-CGAP-SG2 Mus musculus cDNA clone IMAGE:4923619 5'
DEFINITION mRNA sequence.

ACCESSION BG871816
VERSION BG871816.1 GI:14222356
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 1318)
Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

UNPUBLISHED (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
Tissue procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at: <http://image.llnl.gov>

Plate: L1AM10845 Row: d Column: 20
High quality sequence start: 3
High quality sequence stop: 211.
Location/Qualifiers
1. 1318
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4923619"
/lab_host="NCI-CGAP-SG2"
/note="Organ: salivary gland; Vector: pCMV-Sport6; Site: 1; NotI: Site 2; SalI: Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI-CGAP Library."

FEATURES

SOURCE

1. 1318
/organism="Mus musculus"

/db_xref="taxon:10090"
/clone="IMAGE:4923619"
/lab_host="NCI-CGAP-SG2"

/note="Organ: salivary gland; Vector: pCMV-Sport6; Site: 1; NotI: Site 2; SalI: Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 379 a 339 c 297 g 303 t

ORIGIN

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGAGAGCGCCGCACAACTCGAGATTGTATTTTATVN 3']. cDNA prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length b cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAAGAGACTTCCTCAGTTAATAATTAATMCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

75 a 72 C 45 g 107 t

	ch	64.0%	Score 16;	DB 10;	Length 299;
similarity	100.0%	Pred. No. 33;			
16; Conservative	0;	Mismatches	0;	Indels	0; Gaps 0

ggcaagtcacgaact 24
|||||
|||
GGCAAGTCACGAATT 18

BF556014 418 bp mRNA EST 12-DEC-2000
UI-R-A1-dz-d-02-0-UI.r1 UI-R-A1 Rattus norvegicus cDNA clone
UI-R-A1-dz-d-02-0-UI 5', mRNA sequence.
BF556014
BF556014.1 GI:11665744
EST.

Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 418)
Bonaldo,M.F., Lennon,G. And Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

cDNA library preparation: M.B. Soares lab Clone distribution:
This clone is also available through Research Genetics (www.resgen.com)
LLNT (info@llnt.llnl.gov). IMAGE ID= 1771157
Seq primer: M13 Forward.

Location/Qualifiers
1..418
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A1-dz-d-02-0-UI"
/clone_1bp="UI-R-A1"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-A1 library is a subtracted library derived from the UI-R-AO library. The UI-R-AO library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart,

spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-A1) was constructed as follows: PCR amplified cDNA inserts from a pool of approximately 3,840 UI-R-A0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-A0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-A1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."

BASE COUNT 121 a 97 c 117 g 83 t

ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 418;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgcgaagtc 16
|||||

Db 110 GACATTGTCGCAAGTC 95

RESULT 7

LOCUS BG977588 172 bp mRNA EST 12-JUN-2001

DEFINITION PM4-C10084-040101-006-d11 C10084 Homo sapiens cDNA, mRNA sequence.

ACCESSION BG977588

VERSION BG977588.1 GI:14380323

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 172)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,

Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM4&ct=PM4-C10084-

040101-006-d1&cl=2001-01-04&cl=1)

Seq primer: puc18 forward

High quality sequence stop: 165.

FEATURES

source

1. 172

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="C10084"

/dev_stage="Adult"

/note="Organ: colon_lins; Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 37 a 51 c 55 g 28 t 1 others

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 172;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gtgcgaagtcacaga 21
|||||

Db 8 GTGCGAAGTCACAGA 22

RESULT 8

LOCUS AM871559 214 bp mRNA EST 10-MAY-2001

DEFINITION Ta79604.Y1 Bird-Meloiodogme incognita J2 Meloiodogme incognita

cDNA 5' similar to WP:G26D10.4 CE03027; mRNA sequence.

ACCESSION AM871559

VERSION AM871559.1 GI:8005612

KEYWORDS EST.

SOURCE southern root-knot nematode.

ORGANISM Meloiodogme incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;

Tylenchoidea; Heteroderidae; Meloiodogyninae; Meloiodogme.

1 (bases 1 to 214)

McCarter,J., Clifton,S., Chapelli,B., Pape,D., Martin,J., Wyle,T.

, Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.

, Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvilli,R.

, Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Stepte

, M., Allen,M., Person,B., Swaller,T., Harvey,N., Schuck,R., Kohn,S.

, Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and

Wilson,R.

The Washington Univ. Nematode EST Project, 1999

Unpublished (1999)

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

The library was constructed by Uma Rao and David Bird

(david.bird@ncsu.edu) at North Carolina State University. DNA

sequencing by: Washington University Genome Sequencing Center St.

Louis.

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: -40RP from Gibco

High quality sequence stop: 1.

Location/Qualifiers

1. 214

/organism="Meloiodogme incognita"

/db_xref="taxon:6306"

/clone_lib="Bird-Rao Meloiodogme incognita J2"

/dev_stage="enriched for 2nd stage juveniles"

/lab_host="XL0RL"

/note="Vector: ZAP express - PBKCMV (Stratagene); Site_1:

ECORI; Site_2: XhoI; Oligo (dT) primed library. cDNA was

constructed and cloned unidirectionally into the vector

within the 5' EORI and 3' XhoI sites. This library was

constructed by Dr. Uma Rao and Dr. David Bird at North

Carolina state University.

BASE COUNT

59 a 52 c 19 g 84 t

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 214;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatta 25
|||||
Db 184 CAAGTCACAGAAATTA 170

RESULT 9
BE763893 233 bp mRNA EST 19-SEP-2000
LOCUS
DEFINITION RC4-NT0054-120600-015-g05 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE763893
VERSION BE763893.1 GI:10193817
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 233)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=612=RC4-NT0054-120
600-015-g05&t3=2000-06-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 233.

FEATURES

source
Location/Qualifiers
1..233
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0054"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 88 a 27 c 52 g 66 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
|||||
Db 94 CAAGTCACAGAAATTA 108

RESULT 10
BE763892 253 bp mRNA EST 19-SEP-2000
LOCUS

DEFINITION RC4-NT0054-120600-015-g03 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE763892
VERSION BE763892.1 GI:10193816
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 253)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=612=RC4-NT0054-120
600-015-g03&t3=2000-06-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 12
High quality sequence stop: 253.

FEATURES

source
Location/Qualifiers
1..253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0054"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 95 a 31 c 56 g 71 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
|||||
Db 114 CAAGTCACAGAAATTA 128

RESULT 11
AA317110 256 bp mRNA EST 19-APR-1997
LOCUS
DEFINITION AA317110 Lung Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA317110
VERSION AA317110.1 GI:1969448
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 256)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult

, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, M.-C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, J.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodek, A., Gehrm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-T., Marnaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-P., Ferric, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of CDNA sequence

JOURNAL MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)

COMMENT Other-ESTs: TH099418
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
source
Location/Qualifiers
1..256
/organism="Homo sapiens"
/db_xref="ATCC (inhost):117571"
/db_xref="taxon:9606"
/clone_lib="lung"
/dev_stage="adult"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI ; Site_2: XhoI"

BASE COUNT 73 a 57 c 48 g 75 t 3 others

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 256;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
|||||

Db 189 CAAGTCACAGAAATTA 203

RESULT 12
LOCUS 245844 261 bp mRNA EST 14-NOV-1994
DEFINITION HSCVCE041 normalized infant brain cDNA Homo sapiens CDNA clone c-zve04, mRNA sequence.

ACCESSION 245844
VERSION 245844.1 GI:575078
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 261)
Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Foullet, A., Sebastien, K., Kachichis, C. and Tessier, A.
IMAGE: molecular integration of the analysis of the human genome and its expression

JOURNAL MEDLINE C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
COMMENT Contact: Genethon
Genethon-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33-1-772800
Fax: 33-1-778698
Email: genexpress@genethon.fr
Single read.
Genexpress.library_id: C; Genexpress.sequence_id: y1c-zve04
Seq primer: (-21)M13-universal.

FEATURES
source
Location/Qualifiers
1..261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-zve04"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII; Site_2: NotI; sex=Female; dev_stage=3 months old; isolate=muscular atrophy patient; tissue_type=total brain ; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Souares, Psychiatry Dept. Columbia University, USA
Bento Soares, P.N.A.S. in press"

BASE COUNT 77 a 56 c 31 g 97 t

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
|||||

Db 138 CAAGTCACAGAAATTA 124

RESULT 13
LOCUS BB141316 263 bp mRNA EST 28-JUN-2000
DEFINITION BB141316 RIKEN full-length enriched, adult female vagina Mus musculus CDNA clone 9930007004 3', mRNA sequence.

ACCESSION BB141316
VERSION BB141316.1 GI:8796253
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 263)
Konno, H., Aizawa, K., Akahira, S., Akiyama, T., Arahawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, T., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamane, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshitake Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QVat2=QVt-C10100-
231000-457-d12t53=2000-10-23&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 59
High quality sequence stop: 282.
Location/Qualifiers
1..282
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="C10100"
/dev_stage="Adult"
/note="Organ: colon; ins: Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 68 a 75 c 67 g 72 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 282;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 attgtgcagatcac 18
|||||

Db 51 attgtgcagatcac 37

RESULT 16
AA344205 286 bp mRNA EST 21-APR-1997
LOCUS EST50086 Gall bladder t Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION
ACCESSION AA344205
VERSION AA344205.1 GI:1996443
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 286)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Felder,R.A., Bull
C.J., Lee,N.H., Kirschner,E.F., Weinstock,K.G., Gocayne,J.D., White
O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,
Moreno-Palancas,R.F., McDonald,L.A., Naveen,D.T., Pellicino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudke,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utechtack,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimech,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
JOURNAL
TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)

MEDLINE 96026280
COMMENT Other_ESTs: THC98418
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hg1/hg1.html)
Seq primer: M13 Reverse.

FEATURES
Source
1..286
/organism="Homo sapiens"
/db_xref="ATCC (inhost):145738"
/db_xref="taxon:9606"
/clone_lib="Gall bladder I"
/sex="female"
/dev_stage="adult, 25 yrs"
/note="Organ: gall bladder; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
Site_1: EcoRI; Site_2: XhoI

BASE COUNT 83 a 62 c 48 g 91 t 2 others
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 286;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagatta 25
|||||

Db 197 CAAGTCACAGATTATTA 211

RESULT 17
B53227 292 bp DNA GSS 20-JUN-1998
LOCUS CIT-HSP-2009H2.TR CIT-HSP Homo sapiens genomic clone 2009H2, DNA
DEFINITION
ACCESSION B53227
VERSION B53227.1 GI:2607561
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 292)
Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
Unpublished (1997)
JOURNAL
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mda@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
Source
1..292
/organism="Homo sapiens"
/db_xref="GDB:7040812"
/db_xref="taxon:9606"
/clone="2009H2"

/clone.lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelBAC11; site_1: HindIII; site_2: HindIII"

BASE COUNT
ORIGIN

58 a 64 c 90 g 80 t

Query Match 60.0%; Score 15; DB 13; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caatcacagaatta 25
Db 248 CAACTCACAGAAATTA 234

RESULT 18
LOCUS BF743951 317 bp mRNA EST 10-JAN-2001
DEFINITION CM2-BF0826-181000-428-c02 BF0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF743951
VERSION BF743951.1 GI:12070627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 317)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with OKE expressed
sequence tags
JOURNAL proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
COMMENT 20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=CM2&l2=CM2-BF0826-
181000-428-c02&l3=2000-10-18&l4=1)
Seq primer: puc 18 forward
High quality sequence stop: 317.
Location/Qualifiers

FEATURES
source

1..317
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="BF0826"
/dev stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
, 716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT
ORIGIN

96 a 71 c 49 g 101 t

Query Match.

60.0%; Score 15; DB 13; Length 317;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcaagtcacagaatt 24
Db 45 GCAAGTCACAGAAAT 31

RESULT 19
LOCUS BH121038 324 bp DNA GSS 19-JUL-2001
DEFINITION RPCI-24-230E13.TV RPCI-24 Mus musculus genomic clone RPCI-24-230E13
, DNA sequence.
ACCESSION BH121038
VERSION BH121038.1 GI:14964550
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 324)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-230E13.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Sciences
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaod@igf.org

TITLE Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderinfoframe.htm). BAC end
page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 230 row: E column: 13
Seq primer: T7
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..324
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone.lib="RPCI-24-230E13"
/clone.lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; site_1: BamHI; site_2: BamHI;
RPCI-24 Mouse BAC library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT
ORIGIN

93 a 57 c 88 g 86 t

Query Match 60.0%; Score 15; DB 13; Length 324;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcaagtcacagaatt 24
Db 26 GCAAGTCACAGAAAT 12

RESULT 20
LOCUS T30901 327 bp mRNA EST 06-SEP-1995
DEFINITION EST24536 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA

ACCESSION T30901
 VERSION T30901.1 GI:612999
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 327)
 AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulton, R.A., Bult
 C.J., Lee, N., Kirkness, E.F., Weissman, K.G., Gocayne, J.D., White
 O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.-W., Clayton, R.A.,
 Cline, R.T., Cotton, M.D., Batle-Hughes, J., Fine, L.D., Fitzgerald,
 L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S.M., Glodek, A.,
 Gheh, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Kelley, J.M.,
 Kline, K.M., Kelley, J.C., Liu, L.-I., Marmaro, S.M., Merrick, J.M.,
 Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pellegrino, S.M.,
 Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
 Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Wiley,
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
 Dimke, D., Feng, P., Ferrie, A., Fischer, C., Hastings, G.A., He, W.-W.,
 Hu, J.-S., Greene, J.M., Gruber, J., Hudson, P., Kim, A., Kozak, D.L.,
 Kunsch, C., Li, H., Li, H., Melsner, P.S., Olsen, H., Raymond, L., Wei,
 Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon,
 M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
 Venter, J.C.
 TITLE Initial Assessment of Human Gene Diversity and Expression Patterns
 JOURNAL Based upon 83 Million Basepairs of cDNA Sequence
 MEDLINE Nature 377, 3-174 (1995)
 COMMENT Other ESTs: TH010632
 Contact: Venter, J.C.
 The Institute for Genomic Research
 933 Clopper Rd, Gaithersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: tdbinfo@tdb.tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please contact the TIGR database
 (tdbinfo@tdb.tigr.org)
 Seq primer: M13 Reverse.
 FEATURES
 source Location/Qualifiers
 1..327
 /organism="Homo sapiens"
 /db_xref="ATCC (Inhost):100882"
 /db_xref="taxon:9606"
 /clone_lib="Human Brain"
 /note="Organ: brain"
 BASE COUNT 89 a 75 c 47 g 115 t 1 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caagtcacagaatta 25
 ||||||||||||||||
 Db 295 CAAGTCACAGAAATTA 281
 RESULT 21
 BF743964/c 335 bp mRNA EST 10-JAN-2001
 LOCUS DEFINITION CM2-BT0826-181000-428-h04 BT0826 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF743964
 VERSION BF743964.1 GI:12070640
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 335)

AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&tl2=CM2-BT0826-181000-428-h04&ts=2000-10-18&tl=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 40
 High quality sequence stop: 335.
 FEATURES
 source Location/Qualifiers
 1..335
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0826"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2:
 Sma1; A mini-library was made by cloning products derived
 from ORFESTS PCR (U.S. Letters Patent application No. 196
 7716 - Ludwig Institute for Cancer Research) profiles
 into the puc 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."
 BASE COUNT 101 a 76 c 55 g 103 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; Length 335;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 gcaagtcacagaatt 24
 ||||||||||||||||
 Db 61 GCAAGTCACAGAAATT 47
 RESULT 22
 BF743964/c 346 bp mRNA EST 10-JAN-2001
 LOCUS DEFINITION CM2-BT0826-181000-428-d07 BT0826 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF743964
 VERSION BF743964.1 GI:12070630
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 346)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=CW2&t2=CW2-BT0826-18100-428-d07&t3=2000-10-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 43
High quality sequence stop: 346.
Location/Qualifiers

FEATURES

Source

1. 346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0826"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 103 a 78 c 54 g 111 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 346;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtccagaatt 24
|||||
Db 73 GCAAGTCACAGATT 59

RESULT 23
BG466199 349 bp mRNA EST 20-MAR-2001
LOCUS BG466199/c
DEFINITION ELETSED25C05.y1 Elmeria tenella S5-2 cDNA Neg Selected Elmeria tenella cDNA 5' similar to TR:088349 088349 LATENT TGF BETA BINDING PROTEIN.; mRNA sequence.

ACCESSION BG466199.1 GI:13395174
VERSION EST.
KEYWORDS
SOURCE Elmeria tenella.
ORGANISM Elmeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.

REFERENCE 1 (bases 1 to 349)
AUTHORS Liberatori,P., Diaz,C., Tang,K., Marrero,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schenk,R., Rittler,E., Kohn,S., Florence,N., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterson,R., Wilson,R. and Sibley,D.
WashU-Merck Elmeria tenella project
Unpublished (1999)
Contact: David Sibley, Ph.D.
WashU-Merck Elmeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Contact David Sibley (toxoest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 318.

FEATURES

source

Location/Qualifiers
1. 349
/organism="Elmeria tenella"
/strain="1518"
/db_xref="taxon:5802"
/clone_lib="Elmeria tenella S5-2 cDNA Neg Selected"
/dev_stage="Sporozoite stage"
/lab_host="SOLR"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using EXassist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>=10, from 1682 previous reads). Insert sizes range from 1.2-2.9kb. The library may contain a small percentage of host or bacterial contaminants."
BASE COUNT 75 a 107 c 103 g 64 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagt 15
|||||
Db 328 GACATTGTGCGCAAGT 314

RESULT 24
AA648449 353 bp mRNA EST 13-NOV-1997
LOCUS AA648449
DEFINITION ns22f10.s1 NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:1184395 3' similar to contains element LTR5 repetitive element.; mRNA sequence.

ACCESSION AA648449
VERSION AA648449.1 GI:2574878
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 353)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
unknown library type
Insert length: 2027
Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 334.
Location/Qualifiers

FEATURES

source

1. 353
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1184395"
/clone_lib="NCI-CGAP GCBI"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7n3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Altman (NCI) and Dr. Gerald Marli (CBER). cDNA synthesis was

primed with a Not I - oligo(CT) primer
 15'-TGTACCAATTCGAGTCGAGCGCCCTCATTTTCTTTTCTTTTCTTTT-3'
 1. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaudo."

BASE COUNT 131 a 49 c 39 g 134 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||||||||||
 Db 249 CAAGTCACAGAAATTA 263

RESULT 25
 LOCUS T07051 370 bp mRNA EST 30-JUN-1993
 DEFINITION EST004940 Petal brain, Striatogene (cat#936206) Homo sapiens cDNA
 clone HFBC79, mRNA sequence.
 ACCESSION T07051
 VERSION T07051.1 GI:318200
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 370)
 REFERENCE Adams, M.D., Kerlavange, A.R., Fields, C. and Venter, J.C.
 3,400 expressed sequence tags identify diversity of transcripts
 from human brain
 JOURNAL Nature Genet. 4, 256-267 (1993)
 MEDLINE 93364420
 COMMENT Contact: Adams, MD
 The Institute for Genomic Research
 932 Clopper Road, Gaithersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: mdadams@tigr.org
 Seq primer: M13-21.

FEATURES
 source
 1..370
 /organism="Homo sapiens"
 /db_xref="ATCC (lnhost):83698"
 /db_xref="taxon:9606"
 /clone="HFBC79"
 /clone_lib="Petal brain, Striatogene (cat#936206)"
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II
 vector, 1.0kb average insert size."
 BASE COUNT 137 a 58 c 68 g 106 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 370;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||||||||||
 Db 113 CAAGTCACAGAAATTA 127

RESULT 26
 LOCUS A1694429 402 bp mRNA EST 02-JUN-1999
 DEFINITION W83106.x1 NCL-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2338211 3',
 mRNA sequence.

ACCESSION A1694429
 VERSION A1694429.1 GI:4971769
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 402)
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 Tumor Gene Index
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: -400P from Gibco.

FEATURES
 source
 1..402
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2338211"
 /clone_lib="NCI-CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH108"
 /note="Organ: lung; Vector: pT73D-Pac (pharmacia) with a
 modified polylinker; plasmid DNA from the normalized
 library NCI CGAP Lu5 was prepared, and ss circles were
 made in vitro. Following HAP purification, this DNA was
 used as tracer in a subtractive hybridization reaction.
 The driver was PCR-amplified cDNAs from a pool of 5,000
 clones made from the same library (clonoids
 141920-1417991 and 1520904-1522439). Subtraction by Bento
 Soares and M. Fatima Bonaudo."
 BASE COUNT 160 a 60 c 69 g 113 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 402;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||||||||||
 Db 342 CAAGTCACAGAAATTA 356

RESULT 27
 LOCUS A1203847 404 bp mRNA EST 27-JAN-1999
 DEFINITION q175h03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1755893
 3', mRNA sequence.
 ACCESSION A1203847
 VERSION A1203847.1 GI:3756453
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 404)
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaudo

, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
found through the NCI-CCAP clone distribution information can be
www-bio.ln1.gov/bhrp/image/image.html
Insert Length: 563 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 394.
Location/Qualifiers
1..404

FEATURES
SOURCE
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:175893"
/clone_id="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT713D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dt) primer [5'
TGTACCAATCTGAGAGCGAGCGCCGCCAATTTTTTTTTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT713 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bernaldo."
BASE COUNT 161 a 61 c 69 g 113 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatta 25
|||||
Db 342 CAAGTCACAGAAATTA 356

RESULT 28
LOCUS BF726110 405 bp mRNA EST 05-JAN-2001
DEFINITION by01c01.x1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
sapiens cDNA clone by01c01 3', mRNA sequence.
ACCESSION BF726110
VERSION BF726110.1 GI:12042021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 405)
REFERENCE Wistow, G.J., Bernstein, S., Behal, A. and Smith, D.
NEIBANK: EST analysis and bioinformatics for ocular genomics
Invest. Ophthalmol. Vis. Sci. 41 (2000) in press
COMMENT Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 01 row: c column: 01
Seq primer: -21M13 forward primer (AB1).
Location/Qualifiers
1..405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="by01c01"
/clone_id="Human Lens cDNA (Un-normalized, unamplified):
By"
/tissue_type="Lens"

/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life Technologies
, essentially following the protocols of the Superscript
Plasmid System full details of which are contained in the
manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-PGACTAGTTCATGATCGAGCGCCGCC(T)15-3']. Not I/Pbunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT 140 a 58 c 47 g 160 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 405;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatta 25
|||||
Db 249 CAAGTCACAGAAATTA 263

RESULT 29
LOCUS BG413230/6 406 bp mRNA EST 13-MAR-2001
DEFINITION E1ESTed18g03.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
tenella cDNA 5', similar to SW:FBNI_BOVIN P98133 FIBRILIN 1
PRECURSOR, mRNA sequence.
ACCESSION BG413230
VERSION BG413230.1 GI:13318783
KEYWORDS EST.
SOURCE Eimeria tenella.
ORGANISM Eimeria tenella.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
Eimeria.

REFERENCE 1 (bases 1 to 406)
AUTHORS Liberator, P., Diaz, C., Tang, R., Marra, M., Hillier, L., Kucaba, T.,
Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen,
M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey,
N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson,
Y., Cardenas, M., McCann, R., Waterson, R., Willson, R. and Sibley, D.
WashU-Merck Eimeria tenella project
Unpublished (1999)
COMMENT Contact: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Contact David Sibley (toxoest@bcm.tmc.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco.
Location/Qualifiers
1..406
/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_id="Eimeria tenella S5-2 cDNA Neg Selected"
/dev_stage="Sporozoite stage"
/lab_host="SOLR"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Sporozoites were obtained from in vitro sporulated and
excysted oocysts of E. tenella grown in vitro in chickens. cDNA was
synthesized from poly mRNA using an oligo-dT primer
containing a XhoI site. Following second strand synthesis,

ACCESSION A0142435
 VERSION A0142435.1 GI:3533088
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 427)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc.Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3064 row: A column: 21
 Class: BAC ends
 High quality sequence stop: 427.
 FEATURES
 source
 1..427
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Plate-3064 Col=21 Row=A"
 /lab_host="XLR01"
 /note="Vector: ZAP express - PBKCMV (Stratagene); Site:1:
 EcoRI; Site:2: XhoI; Oligo (dT) primed library. cDNA was
 constructed and cloned unidirectionally into the vector
 within the 5' EcoRI and 3' XhoI sites. This library was
 constructed by Dr. Uma Rao and Dr. David Bird at North
 Carolina state University."
 BASE COUNT 137 a 78 c 103 g 108 t 1 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 13; Length 427;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 gcaagtcacagaatt 24
 ||||||||||||
 Db 198 GCAAGTCACAGATT 164
 RESULT 33
 A0870882 429 bp mRNA EST 10-MAY-2001
 LOCUS A0870882/C
 DEFINITION rasi09.yl Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
 cDNA 5' similar to WP:C26D10.4 CE03027 ;, mRNA sequence.
 ACCESSION A0870882
 VERSION A0870882.1 GI:8004935
 KEYWORDS EST.
 SOURCE southern root-knot nematode.
 ORGANISM Meloidogyne incognita
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchima;
 Tylenchoidea; Heteroderidae; Meloidogyinae; Meloidogyne.
 REFERENCE 1 (bases 1 to 429)
 McCarter,J., Clifton,S., Chiappelli,B., Pape,D., Martin,J., Wylie,T.,
 Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
 Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisshvili,R.,
 Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Streptoe
 ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
 Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
 Wilson,R.
 TITLE The Washington Univ. Nematode EST Project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 The library was constructed by Uma Rao and David Bird
 (david.bird@ncsu.edu) at North Carolina State University. DNA
 Sequencing by: Washington University Genome Sequencing Center St.
 Louis.
 Seq primer: -40RP from Glibco
 High quality sequence stop: 428.
 FEATURES
 source
 1..429
 /organism="Meloidogyne incognita"
 /db_xref="taxon:6306"
 /clone_lib="Bird-Rao Meloidogyne incognita J2"
 /dev_stage="enriched for 2nd stage juveniles"
 /lab_host="XLR01"
 /note="Vector: ZAP express - PBKCMV (Stratagene); Site:1:
 EcoRI; Site:2: XhoI; Oligo (dT) primed library. cDNA was
 constructed and cloned unidirectionally into the vector
 within the 5' EcoRI and 3' XhoI sites. This library was
 constructed by Dr. Uma Rao and Dr. David Bird at North
 Carolina state University."
 BASE COUNT 127 a 66 c 38 g 198 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 10; Length 429;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caagtcacagaatta 25
 ||||||||||||
 Db 385 CAAGTCACAGATT 371
 RESULT 34
 A0086771 454 bp DNA GSS 26-AUG-1998
 LOCUS A0086771
 DEFINITION HS-2175_B2_A06_MR CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate-2175 Col-12 Row-B, DNA sequence.
 ACCESSION A0086771
 VERSION A0086771.1 GI:3455988
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 454)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2175 row: B column: 12
 Class: BAC ends
 High quality sequence stop: 454.
 FEATURES
 source
 1..454
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-2175 Col=12 Row=B"

/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pHELOBAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 188 a 70 c 76 g 117 t 3 others

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
|||||

Db 156 CAAGTCACAGAATTA 170

RESULT 35
AA973493 455 bp mRNA EST 17-JUN-1998
LOCUS 0045605.s1 NCI_CGAP_Lu5 Homo sapiens CDNA clone IMAGE:1569152 3',
DEFINITION mRNA sequence.
ACCESSION AA973493
VERSION AA973493.1 GI:3148673
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 455)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA library preparation: M. Bento Soares, Ph.D.
CDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 553 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 440.
Location/Qualifiers
1. 455
FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1569152"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from a
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 175 a 64 c 84 g 132 t

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
|||||

Db 340 CAAGTCACAGAATTA 354

RESULT 36
A1756736 456 bp mRNA EST 18-JAN-2000
LOCUS EESTea24107.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella CDNA
DEFINITION EESTea24107.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella CDNA
5' similar to SW:FBNI_BOVIN P98133 FIBRILLIN 1 PRECURSOR ;, mRNA
sequence.
ACCESSION A1756736
VERSION A1756736.1 GI:5150459
KEYWORDS EST.
SOURCE Eimeria tenella.
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
Eimeria.
REFERENCE 1 (bases 1 to 456)
AUTHORS Liberator, P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,
Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen
M., Bowers, Y., Person, B., Sweller, T., Gibbons, M., Pape, D., Harvey
N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson
Y., Cardenas, M., McCann, R., Waterston, R., Wilson, R. and Sibley, D.
WashU-Merck Eimeria tenella project
UNPUBLISHED (1999)
CONTACT: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Contact David Sibley (toxoe@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40R from Gibco
High quality sequence stop: 419.
Location/Qualifiers
1. 456
FEATURES
source
/organism="Eimeria tenella"
/strain="IS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria S5-2 Sporozoite stage"
/dev_stage="Sporozoite"
/lab_host="SOLR E. coli"
/note="vector: Bluescript SK-; Site.1: EcoRI; Site.2: XhoI
; Sporozoites were obtained from in vitro sporulated and
excysted oocysts of E. tenella grown in chickens. cDNA
was synthesized from poly mRNA using an oligo-dT primer
containing a XhoI site. Following second strand synthesis,
EcoRI adaptors were ligated to the cDNA and products were
size-selected on Sephacryl S500. cDNAs were digested with
EcoRI/XhoI and cloned into lambda zap II (Stratagene).
Clones were converted to phagemids by mass excision using
Exassist helper phage and SOLR cells (Stratagene).
Insert sizes range from 1.2-2.9 kb."

BASE COUNT 100 a 137 c 132 g 87 t

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacattgtcgaagt 15
|||||

Db 435 GACATTGTCGCAAGT 421

RESULT 37
A1361386 457 bp mRNA EST 15-FEB-1999
LOCUS qy47f01.x1 NCI_CGAP_Brn23 Homo sapiens CDNA clone IMAGE:2015161 3',
DEFINITION mRNA sequence.
ACCESSION A1361386

VERSION AT361386.1 GI:4113007
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 457)
 AUTORS NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute / National Institute of Neurological
 Disorders and Stroke, Brain Tumor Genome Anatomy Project
 (CGAP/BRGAP), Tumor Gene Index
 JOURNAL Unpublished (1998)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
 Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Insert length: 1228 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 434.
 Location/Qualifiers
 1..457
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:2015161"
 /clone_1ib="NCI-CGAP_Brn23"
 /tissue_type="glioblastoma (pooled)"
 /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5',
 TCGTACCATCTGCACTGGAGCGCCGATCTCTTTTCTTTTCTTTTCTTTTCTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT7T3D vector.
 Library is normalized, and was constructed by Bento
 Soares and M.Fatima Bonaldo."
 BASE COUNT 153 a 73 c 52 g 179 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 10; Length 457;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caagtcacagaatta 25
 Db 153 CAAGTCACAGAATTA 167
 RESULT 38
 BG562271 464 bp mRNA EST 10-APR-2001
 LOCUS BG562271/c
 DEFINITION Ectestge05a02.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
 tenella cDNA 5', similar to SW:FBNI-BOVIN P98133 FIBRILLIN 1
 PRECURSOR; mRNA sequence.
 ACCESSION BG562271
 VERSION BG562271.1 GI:13591269
 KEYWORDS EST.
 SOURCE Eimeria tenella.
 ORGANISM Eimeria tenella.
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
 Eimeria.
 REFERENCE 1 (bases 1 to 464)
 AUTHORS Liberatore,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
 Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
 M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey

TITLE 'N', Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
 Y., Cardenas,M., McCann,R., Waterson,R., Wilson,R. and Sibley,D.
 JOURNAL WashU-Merck Eimeria tenella project
 COMMENT Unpublished (1999)
 Contact: David Sibley, Ph.D.
 WashU-Merck Eimeria tenella project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Contact David Sibley (toxoest@borcim.wustl.edu) for further
 information relating to organism, libraries, or clone availability.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 424.
 Location/Qualifiers
 1..464
 /organism="Eimeria tenella"
 /strain="LS18"
 /db_xref="taxon:5802"
 /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
 /dev_stage="Sporozoite stage"
 /lab_host="SOLR"
 /note="Vector: Bluescript SK-; Site:1: EcoRI; Site:2: XhoI
 ; Sporozoites were obtained from in vitro sporulated and
 excysted oocysts of E.tenella grown in chickens. cDNA was
 synthesized from poly mRNA using an oligo-dT primer
 containing a XhoI site. Following second strand synthesis,
 EcoRI adaptors were ligated to the cDNA and products were
 size-selected on sephacryl S500. The cDNA were ligated to
 EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
 converted to phagemids by mass excision using Exassist
 helper phage and E.coli SOLR cell (Stratagene). Clones
 were selected by negative hybridization against a pool of
 overrepresented ESTs (N>=10, from 1682 previous reads).
 Insert sizes range from 1.2-2.9kb. The library may contain
 a small percentage of host or bacterial contaminants."
 BASE COUNT 101 a 141 c 137 g 85 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; Length 464;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gacattgcgaagt 15
 Db 421 GACATTGCCAAGT 407
 RESULT 39
 A1635259 471 bp mRNA EST 16-DEC-1999
 LOCUS A1635259
 DEFINITION t279h07.x1 NCI-CGAP_Paul Homo sapiens cDNA clone IMAGE:2294845 3',
 mRNA sequence.
 ACCESSION A1635259
 VERSION A1635259.1 GI:4686589
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 471)
 AUTORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Life Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1533 Std Error: 0.00
 Seq primer: -400p from Gibco
 High quality sequence stop: 399.
 Location/Qualifiers

FEATURES
 source
 1..471
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2294845"
 /clone_lib="NCI_CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: PCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

BASE COUNT 171 a 69 c 53 g 178 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 471;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatt 25
 |||||||
 Db 235 CAAGTCACAGAAAT 249

RESULT 40
 LOCUS BG020076 485 bp mRNA EST 24-JAN-2001
 DEFINITION dc47c07.x1 NICHG XGC Emb3 Xenopus laevis cDNA clone IMAGE:3400308
 3' similar to TR:Q9VCA6 Q9VCA6 G6313 PROTEIN.; mRNA sequence.
 ACCESSION BG020076
 VERSION BG020076.1 GI:12476155
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 485)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbrr@mail.nih.gov
 Tissue Procurement: Martha Rehbert, Steven L. Klein, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNI)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/LLNI at: info@image.llnl.gov
 Seq primer: -400p from Gibco
 High quality sequence stop: 409.
 Location/Qualifiers

FEATURES
 source
 1..485
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:3400308"
 /clone_lib="NICHG XGC Emb3"
 /tissue_type="embryo (stages 24-25)"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 1.7 kb. Constructed by Life Technologies. Note: This
 is a Xenopus Gene Collection (XGC) library."

BASE COUNT 117 a 102 c 91 g 165 t 10 others
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 485;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatt 24
 |||||||
 Db 437 GCAAGTCACAGAAAT 451

RESULT 41
 LOCUS BG516037/c 487 bp mRNA EST 30-MAR-2001
 DEFINITION E15E9d8c02.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
 tenella cDNA 5' similar to SW:PBNI_BOVIN P98133 FIBRILLIN 1
 PRECURSOR.; mRNA sequence.
 ACCESSION BG516037
 VERSION BG516037.1 GI:13489363
 KEYWORDS EST.
 SOURCE Eimeria tenella.
 ORGANISM Eimeria tenella
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;

REFERENCE 1 (bases 1 to 487)
 AUTHOR Liberators, P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,
 Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen
 M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey
 N., Schurr, R., Rittler, E., Kohn, S., Florence, N., Shin, T., Jackson
 Y., Cardenas, M., McCann, R., Waterston, R., Wilson, R. and Sibley, D.
 WashU-Merck Eimeria tenella project
 Unpublished (1999)
 JOURNAL Contact: David Sibley, Ph.D.
 COMMENT WashU-Merck Eimeria tenella project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Contact David Sibley (toxoest@borcim.wustl.edu) for further
 information relating to organism, libraries, or clone availability.
 Seq primer: -400p from Gibco
 High quality sequence stop: 403.
 Location/Qualifiers

FEATURES
 source
 1..487
 /organism="Eimeria tenella"
 /strain="PS18"
 /db_xref="taxon:5802"
 /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
 /dev_stage="sporozoite stage"
 /lab_host="SOLR"
 /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
 ; Sporozoites were obtained from in vitro sporulated and
 excysted oocysts of E. tenella grown in chickens. cDNA was
 synthesized from poly mRNA using an oligo-dT primer
 containing a XhoI site. Following second strand synthesis,
 EcoRI adapters were ligated to the cDNA and products were
 size-selected on sephacryl S500. The cDNA were ligated to
 EcoRI/XhoI prepared lambda zapII (stratagene). Clones were
 converted to phagemids by mass excision using Exsist
 helper phage and E.coli SOLR cell (Stratagene). Clones
 were selected by negative hybridization against a pool of
 overrepresented ESTs (N>10, from 1682 previous reads).
 Insert sizes range from 1.2-2.9kb. The library may contain
 a small percentage of host or bacterial contaminants."

BASE COUNT 105 a 147 c 142 g 93 t
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 487;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacattgtcagaatt 15
 |||||||

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 502)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapsb@mail.nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 , Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html
 Insert length: 925 Std Error: 0.00
 Seq primer: -40ml3 fwd. EJ from Amersham
 High quality sequence stop: 448.
 Location/Qualifiers
 1..502
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1394285"
 /clone_id="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - Oligo(dT) primer [5',
 TGTTACCAATCTGACAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 196 a 67 c 97 g 142 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 502;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||
 Db 342 CAAGTCACAGAAATTA 356

RESULT 45
 A1672463 502 bp mRNA EST 16-DEC-1999
 LOCUS wa03605.x1 NCL_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2297024 3',
 DEFINITION mRNA sequence.
 ACCESSION A1672463
 VERSION A1672463.1 GI:4852194
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 502)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapsb@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.

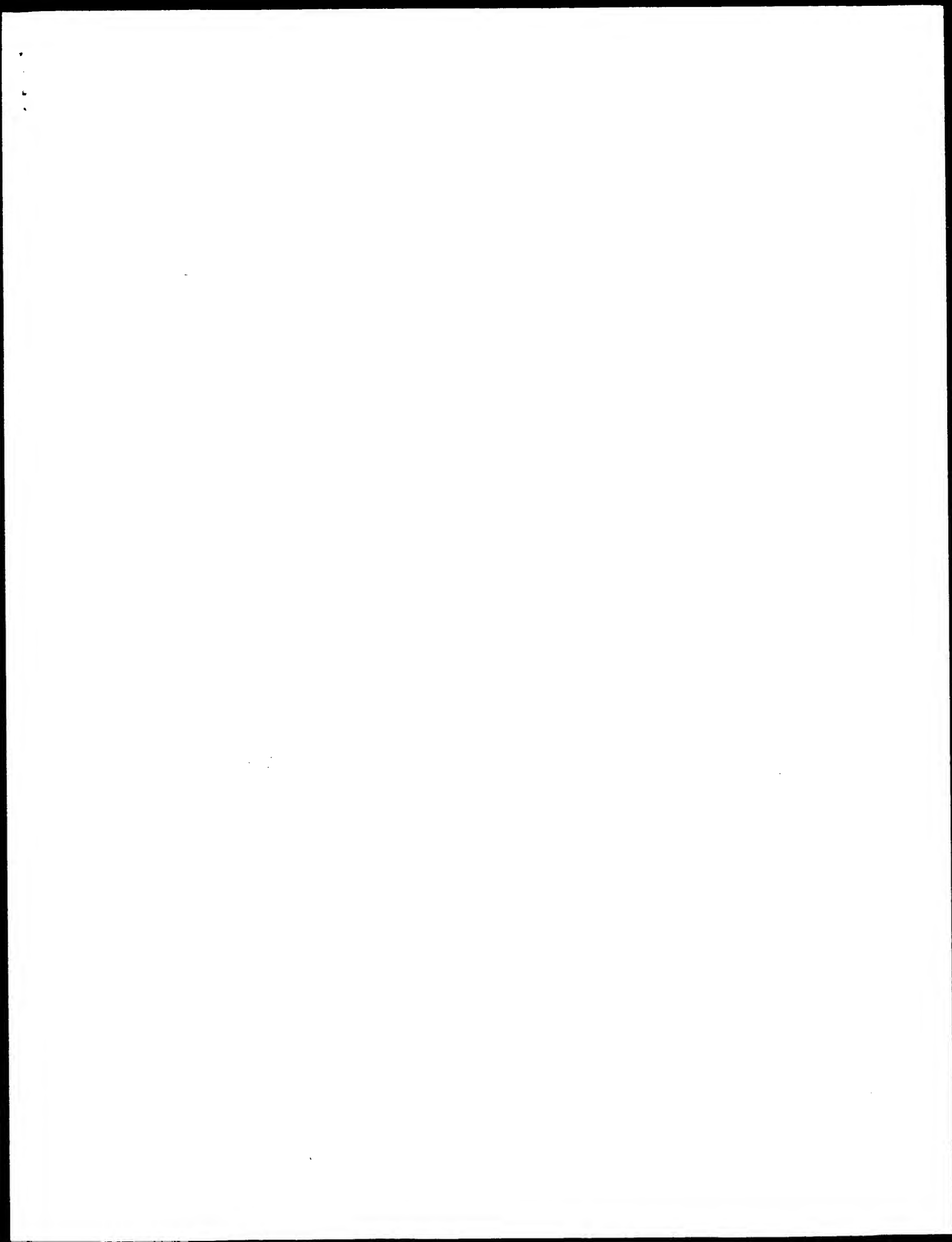
CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html
 Insert length: 535 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 458.
 Location/Qualifiers
 1..502
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2297024"
 /clone_id="NCL_CGAP_Kid11"
 /lab_host="DH10B"
 /note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker. Site_1: Not I; Site_2: Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (cloneids 1322376-1323911, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldo."

BASE COUNT 198 a 67 c 96 g 141 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 502;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||
 Db 339 CAAGTCACAGAAATTA 353

Search completed: December 26, 2001, 14:09:53
 Job time: 7040 sec



71 15 60.0 4674 9 AF034373 Homo sapi
72 15 60.0 5526 1 AF250776 Homo sapi
73 15 60.0 8482 9 AF281036 Homo sapi
74 15 60.0 12891 1 AE004192 Homo sapi
75 15 60.0 15143 9 HSAJ4862 Homo sapi
76 15 60.0 20780 2 AC014847 Homo sapi
77 15 60.0 22773 9 AF107890 Homo sapi
78 15 60.0 23185 2 AC014149 Homo sapi
79 15 60.0 32886 2 HSU201H11 Human DNA
80 15 60.0 39017 9 HSJ850N15 Human DNA
81 15 60.0 44706 2 AC022328 Mus muscu
82 15 60.0 50149 9 AC011955 Homo sapi
83 15 60.0 60100 9 AL138961 Human DNA
84 15 60.0 61170 2 AC068163 Homo sapi
85 15 60.0 68334 2 AC010563 Homo sapi
86 15 60.0 73167 2 AC010316 Homo sapi
87 15 60.0 79288 3 AE003572 Drosophi
88 15 60.0 79938 2 AC025399 Homo sapi
89 15 60.0 87417 2 AL389896 Homo sapi
90 15 60.0 92588 9 HSIIMDL X97051 DNA sequenc
91 15 60.0 102326 9 AC087407 Homo sapi
92 15 60.0 106482 10 AC007585 Mus muscu
93 15 60.0 110000 2 AC087782-2 Continuation (3 of
94 15 60.0 110000 2 AL591074-4 Continuation (5 of
95 15 60.0 110000 2 LMFLCHR31-11 Continuation (12 of
96 15 60.0 118561 2 AC009217 Drosophi
97 15 60.0 122914 10 AC069451 AC006941 Mus muscu
98 15 60.0 136804 9 AP000469 Homo sapi
99 15 60.0 137889 9 AC073269 Homo sapi
100 15 60.0 141704 9 AC011754 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
Source Location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25
|||||
Db 111 TCCTCAGGTGCAGGTGCAGCAGCTTG 135

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
Source Location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25
|||||
Db 111 TCCTCAGGTGCAGGTGCAGCAGCTTG 135

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
JOURNAL Escherichia.
FEATURES
Source Location/Qualifiers
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/gene="B10B"
24..1064
/gene="B10B"
/codon_start=1
/transl_table=11
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
/translation="MAHRPRRTLSQVTELEFERPLDLLFEAOQVHROHDPPOVOST
LSTKTCACEPDCKYCCROSSRYRTGLAEERLMEBOVLESARAKAGSTRCMCAAW
KNPHERDMPYLEQWQVKAMGLEACHTTGLTSSQAORLANAGLDYVHNNDLSPER
YGNITTRTYOERLDLLEKVRDAGIKVSGGIVGLGETVDRAGLLQLANLPTPEP
VPINMLKVKVGTPLADNDVDADDFIRITAVARIMPTSYVRISAGREOMDOTAMC
FMAGANSIFPGCKLITTPNPEEDKDLQLEFRKGLINQQAVALAGNDEQQRLEQALMT
PDDEYVNAAL"

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25
|||||
Db 134 TCCTCAGGTGCAGGTGCAGCAGCTTG 158

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin synthetase.
ACCESSION E00893
VERSION E00893.1
KEYWORDS
SOURCE

```

ACCESSION   E00893
VERSION     GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 1121)
AUTHORS    Hirono, Y., Kojima, T. and Kimura, H.
TITLE      DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
JOURNAL     Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS Escherichia coli
            PN JP 1986149091-A/1
            PE 07-JUL-1986
            PI 24-DEC-1984 JP 1984272605
            PI HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC strandedness: Double;
            CC topology: Linear;
            CC hypothetical: No;
            CC anti-sense: No;
            CC *source: strain=Escherichia coli Ns101;
            CC feature is identified by experimental;
            FH Key Location/Qualifiers
            FT CDS 42..1079
            FT /product='biotin synthetase'.
FEATURES    source
            location/Qualifiers
            1..1121
            /organism="Escherichia coli"
            /db_xref="taxon:562"
BASE COUNT  289 a 296 c 325 g 211 t
ORIGIN
Query Match 100.0% Score 25: DB 6: Length 1121:
Best Local Similarity 100.0%: Pred. No. 0.0027:
Matches 25: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 1 tcgtcaggtcagcagtcagcttg 25
    |||||||||||||||||||||||
Db 152 tcgtcaggtcagcagtcagccttg 176

RESULT 5
ECOBIO      5793 bp DNA BCT 28-FEB-1994
LOCUS       E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
DEFINITION (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc
            protein, and dehydrobiotin synthetase (bioD), complete cds.
ACCESSION   J04423.1 GI:145422
VERSION     7,8-diamino-pelargonic acid aminotransferase;
KEYWORDS    7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioc gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
ORGANISM    Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5793)
AUTHORS    Otsuka, A.J., Buoncrisiani, M.R., Howard, P.K., Flamm, J. and
            Johnson, O.
TITLE      The Escherichia coli biotin biosynthetic enzyme sequences
JOURNAL     J. Biol. Chem. 263, 19577-19585 (1988)
COMMENT     Predicted
            Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES    location/Qualifiers
            gene
            CDS
            1..5793
            /organism="Escherichia coli"
            /strain="K-12"
            /db_xref="taxon:562"
            complement(98..574)
            /note="orf 1"
            /codon_start=1
            /transl_table=1
            /protein_id="AAA23513.1"
            /db_xref="GI:455168"
            /translation="MKLISNDLRDGDCLKPHRHVNGKGYDDGDNISPHLADVPAGTK
            SEVVYCDPDAPFGSGMHWVNLPAIDRVLPQFGSGGLVAMPDGLVLOTDTGKTS
            YDGAAPKKGERTHYFTVHALDIERIDVDEGASAMVGFNVHFLSASITAFMS"
            complement(633..1925)
            /gene="bioA"
            /complement(633..1925)
            /gene="bioA"
            /codon_start=1
            /transl_table=1
            /product="7,8-diamino-pelargonic acid aminotransferase"
            /protein_id="AAA23514.1"
            /db_xref="GI:457106"
            /translation="MTTDLAPDQRIHPTYSMTSPLPYVYVSAEGCELLISGR
            LVDGMSWMAAIGHVNHQNLNAAKMSQIDAMSHVFGGTTAPATIELCRKLVAMGRN
            ALECYFLADSGVAAYEVAMKMALOYWOAKGERORFLFRNGYHGDITFGAMSCDPN
            SMHSLMKGYLPENLFAPAPQSRMDGEMDERDVGARLMAARHBIAAVITIEPYGA
            GGRMWHPEMLKRIKICDRREGILIADEIATGFGRTGFLACAAASIALIIESGDM
            LNCMTLSATLTITTEVAETISNGEACGPMGTGNPLACAAASIALIIESGDM
            OOVADIEVOLREQLAPARDAENVADYRVIGATGVETHTPVMAALQPFVGGVWIR
            PEKLLIYMPYIILFPOLQRLTAAYNRAVDTEFFCO"
            2012..3052
            /gene="bioB"
            2012..3052
            /gene="bioB"
            /codon_start=1
            /transl_table=1
            /product="biotin synthetase"
            /protein_id="AAA23515.1"
            /db_xref="GI:145425"
            /translation="MAHRPMTLSQYTELEFEKPLDLLFEAOQVHRHEDPPOVST
            LISIKGACPEDECKYCPOTSRYKTLGEARIMEVEQVLESAKRAKAGSTRCMGAM
            KNFHERDMPLEOVGVKRAMGLEACMTICTISSEAOPLAAGIDYVHNHIDTSEF
            VGINITTRYQERLDLIEKVRDAGIKVCSGIVGLDETYKDRAGLLQLALPTPEPS
            VPINMLVKVGTPLADNDVDADPFIKTAVARIMPTSYVRLSAGREDBMSNOTAMC
            FMAGANSIFYGCKLITTPNEEDBKDLFRKILGNLPOQYAVLAGDNEOQRLQALMT
            PDNDEYXNAAL"
            3049..4203
            /gene="bioF"
            3049..4203
            /gene="bioF"
            /codon_start=1
            /transl_table=1
            /product="7-keto-8-amino pelargonic acid synthetase"
            /protein_id="AAA23516.1"
            /db_xref="GI:145426"
            /translation="MSWQEKINALDARRAADALRRRYPVAGAGMIVADROLYNF
            SSNDYILSHHPQIIRAMQOAGQFGSGSGSVHGVHDALEELAEWLGYSRA
            LFTSGFAANOAVIAAMAKEDRIADRSHASILSAASLSPQLRFPANDHTILAR
            LIAFPCPGQOMVTEGEVFSMDGSAPIAETIOQYTOOHMGMLVDADAGTGVTGEORG
            SLMQOKVPELVITTEKGFSGGAIVICSTVADYILOFARHLIYSTSPPOAOAL
            RASLAVIRSBEGDARRKIALILTRPRAGVODLPFTLADSCSAIOPLIYNGSRAOL
            AEKRUQGCQWTAIRPPTVAGTARLRLTLTAHHEMODIDRLLEVLHGNC"
            4190..4945
            /gene="bioc"
            4190..4945
            /gene="bioc"
            /codon_start=1
            /transl_table=1
            /protein_id="AAA23517.1"
            /db_xref="GI:145427"
            /translation="MATVKNQVIAAAGFGRAAHYEGHADILQROSADALLAMLPRKYT
            HVIDAGCGPMGMSHMRERHNAOVYALDISPPMIYQAKORAAOHVLAGDIESPLATA
            TFDLAWSNLAVQKGNLSTALRELIRYVRKGVAFPTLVGSLPERHQAQVADDERP

```

```

HNRFLPDEIFOSINGVHYHHIOPITLTFEDALSNMSLKIGATLHHEGRPRIL
TRSOLOPLOLAMPQOQGRPLTYHLFLGLVIARE
4938..5397
/gene="b10d"
/db_xref="GI:2294845"
/feature="b10d"
/codon_start=1
/transl_table=11
/product="gelatinase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVITDTEVEGKTAVSCALLOAKRAGYRTAGYKPVASGS
EKTPGRLNSDALOIRNSLOLDYAVNVPYPAETSPHII SAOGRPEIESIYASAG
LRLEOQADWLVGAGGMPETPLSDTFTADWYQPOLPVILVYKKEGIRHMLTA
OYTOHAGLTLAGVANDVTPPKRAHEVMTTLTRMLPRCWEKSPGLQIKRQPESES
T"
BASE COUNT      1363 a 1554 c 1631 g 1245 t
ORIGIN           4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 1; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 tgcacagtgcaagtcagcacgttg 25
        |||||||||||||||||||
Db      2122 TCGTCAGTGCAGTGCAGTGCAGCTTG 2146

RESULT  6
LOCUS   A38246      5872 bp      DNA
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION   A38246.1 GI:2294844
KEYWORDS
SOURCE    Escherichia coli.
           Escherichia coli.
           Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS   Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL    Patent: WO 9408023-A 1 14-Apr-1994;
           LONZA AG (CH)
COMMENT   Other publication PL 308301 950724
           Other publication CA 2145400 940414
           Other publication AU 4820293 940426
           Other publication HU 71781 960228
           Other publication SK 42095 951108
           Other publication CZ 9500809 950913
           Other publication FI 951547 950331
           Other publication JP 8501694T 960227.
FEATURES
source     1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PRO30A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..28
            /standard_name="PROMOTER PTAC"
            /evidence=experimental
            45..49
            /standard_name="B10B RBS NO.9"
            /evidence=experimental
            117..1157
            /gene="B10B"
            117..1157
            /gene="B10B"
            /gene="B10B"

gene
CDS

```

stem_loop 5583. .5605
terminator 5583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcgtcaggtgcaggtcagcagctg 25
|||||
db 227 TCGTCAGGTGCAGGTGCAGCAGCTTG 251

RESULT 7
LOCUS A38251 5872 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)

COMMENT

Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
Location/Qualifiers

FEATURES

source 1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1134..2308
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"

RBS
gene
CDS
1134..2308
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1134..2308
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"

RBS
gene
3030..3045
/standard_name="B10D RBS"
3043..3753
/gene="B10D"

CDS

3043..3753
/gene="B10D"
/standard_name="DEPHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DIB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcgtcaggtgcaggtcagcagctg 25
|||||
db 227 TCGTCAGGTGCAGGTGCAGCAGCTTG 251

RESULT 8
LOCUS A93674 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)

FEATURES
source 1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/function="PROMOTER PRAC"
/evidence=experimental
23..28
/standard_name="PROMOTER PRAC"
/evidence=experimental
105..119
/standard_name="B10B RBS NO.9"
/evidence=experimental
117..1157
/gene="B10B"
117..1157
/gene="B10B"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="B10F1N SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"

RBS
gene
CDS
-35..signal
-10..signal
117..1157
/standard_name="B10B RBS NO.9"
/evidence=experimental
117..1157
/gene="B10B"
117..1157
/gene="B10B"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="B10F1N SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPWTLSQVTELEFEKPLDLLPFAQOVHROHPPROVOYST
LLSTKTGACPDCKYCCPOSSRYKTKGLEAERLMEVEQVLESARKAKAAGSTRFCMAAM

KNHEEDMPYLEDWGVKAMGLBACHTLCTLSQAQRLANAGLDYVNHNDLTPSP
YGLNITRTYOERLDLEKVDAGIKVCSGIVGLSTVAGLLELANLPTPES
VPIIMLVKVGITPLADNDVADPFIETIIVARIMPTSVRLSAGREOMNEOTAMC
PMAGANSIFYGCKLITTPNPEEDKDLQFLKLCILNPQQAIVLAGNEQQRLEQALMT
PMDVEYNAAL"
2284. .2297
/standard_name="BIOC RBS"
2295. .3050
/gene="BIOC"
2295. .3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVNRKQAIANAFAAGAAHYEQHADLQROSADALLAMIPORRYT
HYLDAGCGPMNRHREHRAOYALDISPPIVQAKDADHILADIESPLATA
TFDLASNLAVOMCGNLSPALRELYVVPKGVAFETTLVQSLPELHOAWAVDERP
HARRELPPEIDBSLNGVHYQHIIQPTTLFPDALSAMRSILKIGATHLHEGRDPRIL
TRQRLQRLQAWPQDQGRYPLTYHLPLGVIAFE"
3742. .3752
/standard_name="BIOA RBS"
3750. .5039
/gene="BIOA"
3750. .5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAEFDORHIMHPTSMTSPLVYVNVASGECCELLSDGR
LYDGMSSWMAIIGHYNNPOLNAAKMSQIDAMSHVMFGGITHAPALECRKIYAMTPR
LECVFLADGSVAEVAAMKALOWMAGSEARORFLTRNGYHGTFCAMSVCDPDS
MHSILKGYLPENLFAPAROSRMDGEMDERMDVGFARLMAAHRHEIAAVIIEPTVQAG
GMRMYHPELAKIRIKICDEBGLLIADELATGFKTKLFACEHAELIADILCLGKAL
TGGITLSTLITREVALETISNGEAGCFMHGTEGNGNPLACAAANASIALLESDDMOQ
OVADIEVOLREQLARDAENWADVRLGALGVETTHPVNMAALOKKFEVQGWITRP
FKLTIYMPYIILPQOLRLIAVNRVAVQETPCQ"
5088. .5100
/standard_name="ORFI RBS"
5098. .5574
/gene="ORFI"
5098. .5574
/gene="ORFI"
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDQKLPBHFVNGMGYDDNISPLANDDVPAKIK
SEVYCYDPPADPTGSGMWHYVNLPRADTRILPQCGSLVAMPDGLQTRTDGKIG
YDGAAPKGEHRITFYHALDIRIDVDSAGAVGVNHFHSLASASITAFMS"
5583. .5603
5583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgcaggtcgaagtcagcagcttg 25
|||||
Db 227 tgcgcaggtcgaagtcagcagcttg 251
RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
REFERENCE
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-Oct-1997;
LONZA AG (CH)
FEATURES
source
location/qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141. .1156
/standard_name="BIOF RBS"
1154. .2308
/gene="BIOF"
1154. .2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSMOEKINALDARRARADALRRRYVYAGAGRMVLADROYLNF
LFTSGRAQADVAITAMAKEDRIADRLSHASLEASLSQRLRFANHNVTHLAR
LIASPCQGMVYTEGVFSMDGDSAPLAEIQVTOQNGMVLWVDHAGTGVGEQGRG
SCWLQVQKPELIVTFKGVGSGAVALCSSTVADYLIQFAHLLYSTMPRAQOAL
RASIAVIRSDGADAREKLEALITFRACGVODLPFLADSCSAIOLPIYIGDMSRALOL
AEKLRQCGCWTAIRPVPAGTARLRLTLTAHHEMODIDRLLEPLHNG"
3030. .3045
/standard_name="BIOD RBS"
3043. .3753
/gene="BIOD"
3043. .3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRIYVGTDFEVKIVASCALLQAAKAGYRTAGYKRPVYASG
EKTPGRLNSDALAQRNSLQDLVATVNPYTFAPRTPHIIISAQEGRIEISLVMSAG
LRALQDQDWVLYVEGAGGMPFLSDPTFFADWVTOEQLVILVGVKIGCIINHMLTA
QVIOHAGLLIAGWVANDVTPPGKRAHEVMTTLTRIPARLLGELPWLAEENENATKQ
YINLAFVASTLIGFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.002;

Matches	25;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	tgcTcaggtgcagTccacgacttg	25						
Dd	227	TGTCAGGTGCACGTCACACCTTG	251						

Oy	1	tgcgcaggtgacgttcagcacgttg	25
Dd	227	TGTCAGGTGCAGGTCACGCCCTTG	251

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AR101809	Sequence 1 from patent US 6083712.	AR101809	AR101809.1	GI:12812607	Unknown.	Unknown.

BASE COUNT	1318 a	1552 c	1695 g	1307 t
ORIGIN	/organism="unknown"			

Query Match	100.0%;	Score 25;	DB 6;	Length 5872;
Best Local Similarity	100.0%;	Pred. No. 0.002;		
Matches	25;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0

Oy	1	tccgtcaggtgcaggttcagcacglttg	25
Dd	227	TCGTCAGGTGCAGGTCAAGCAGCTTG	25

RESULT 11		14-FEB-2007	
ARI01810			
LOCUS	ARI01810	5672 bp	DNA
DEFINITION	Sequence 6 from patent US 6083712.		
ACCESSION	ARI01810		
VERSION	ARI01810.1	GI:12812608	

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	source
UNKNOWN:	UNKNOWN:	1 (bases 1 to 5872)	Bitch,O., Brass,J., Fuhrmann,M. and Shaw,N	Biotechnological method of producing biotin	Patent: US 6083712-A 6 04-JUL-2000;	Location/Qualifiers	
						1..5872	
BASE COUNT		1318 a	1552 c	1695 g	1307 t		/organism="unknown"
ORIGIN							

Query Match	100.0%	Score 25;	DB 6;	Length 5872;
Best Local Similarity	100.0%	Pred. No. 0.002;		
Matches 25; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

QY	1	tgcgtcaggtgcaagtcacgaagt	25
Dd	227	TGCTCAGGTGCAGGTCACGACGTTG	25

RESULT	12		
AE000180			
LOCUS	AE000180	11022 bp	DNA
DEFINITION	Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.		
ACCESSION	AE000180	U00096	
			01-DEC-2000

VERSION	AE000180.1	GI:1786988
KEYWORDS	.	
SOURCE	Escherichia coli K12.	
COMMENTARY	Escherichia coli K12	

REFERENCE	AUTHORS	TITLE
1 (Pages 1 to 1102)	Blattner, F. R., Plunkett, G. III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y.	The complete genome sequence of <i>Escherichia coli</i> K-12
Science 277 (5331), 1453-1474 (1997)		
97426617		

FEATURES

SOURCE

repeat_region

gene

CDS
 complement(147..1430)
 /gene="ybhC"
 /function="orf; Not classified"
 /note="f427: 98 pct identical to fragment YBHC_ECOLI
 SW:P46130 (300 aa) but contains 127 additional C-terminal
 residues"
 /codon_start=1
 /transl_table=11
 /product="putative peptidase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTFVSRLALAFGVLTACSTPPDQPSDQTAAPGSRPI
 LSAEADNEDAOHYFASLTGCAAMNPSPITIPADPPVGPAGCPGCTHTTTQAQV
 AAIKTRKROXYLAIVMGEYGVYVPAPAGGTTIGTGKPLDVAIGLSLQSGMSPA
 DMRHDVPRKRYMPKPARMYDSCSKSDSITGVCSAFVFSQNNGLQNLQNTENT
 LGDSVADGNHFAVALRTDGDVOJINNVNLLGRONTFVNSGVNRLFTNRQPLTAV
 NSTYEGVDIVSGRGAIVEDNTEFRVYNSRTQOEAFAATLSNIYGFLLVNSFRN
 AFEDGVAQLGRSLDVADANTNGQVYIRDSAINEGFNAPKPMADAVISNRPFGNTGSVD
 DNDEIQNLNDIYNNRMMEYNNRNGVSKVAEAKK"
 complement(1507..1535)
 /note="factor Sigma70: predicted +1 start at 806574"
 complement(1582..2058)
 /gene="ybhB"
 /note="b0773"
 complement(1582..2058)
 /function="orf; Unknown"
 /note="f158: 99 pct identical to YBHB_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRDDPKLPHRHVFNMGYGDGNI SPLAMDVPAQIK
 SFVNYDDPAPGSGMHWVYVNLPAQTRVLPQFGSGSLVAMPDGVLTQTRDGRGTG
 YGCAAPKGETHRVYITVALIERIDVDEGASGAMVGFVHFSLASSTVAMS"
 complement(2108..2124)
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rhas predicted site"
 complement(2117..3406)
 /gene="b10A"
 /note="b0774"
 complement(2117..3406)
 /gene="b10A"
 /EC_number="2.6.1.62"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="f429: 100 pct identical to B10A_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTDIDIAFDQHIWHPYTSMTSPVPPVSAECCILTSDGRK
 LVDGSSMAAIGHYNNHNPOLNAAKMSQIDASHVYFGTITAPALEICRKLVAATPOP
 LECVYLDSGSAVAVAKMALDTYQAGGEARQRLFRNCHDDETCAMSVCDPDS
 MHSIMKGTLPENLFAPQSRMDGEMDERDVGARLMAARHETAAVITLEPIVQAG
 GMRVYHPEMLKRIKICDRGILLIADIEATGRTGLFACHEAETAPDLICGKAL
 TGGTMTLSATLTTRVATETISNGEACGFMGPTMGNTLACAAANASLALIESGDMQ
 OVADIEVOLREGIAPARDAENVADVRYGATGVETTPVMAALQKFEVQGVATRP
 FGKLIYMPYIILPQOLRLTAANRNVQDETFFCQ"
 complement(2193..2221)
 /gene="b10A"
 /note="factor Sigma70: predicted +1 start at 807260"
 3411..3450
 /note="central position to bioB promoter: -20"
 /bound_moiety="BioB documented site"
 3411..3450
 /note="central position to predicted promoter: -20; genetic
 evidence for the site"
 /bound_moiety="BioB predicted site"
 complement(3411..3450)
 /note="central position to predicted promoter:50"

protein_bind
 complement(3411..3450)
 /note="central position to bioA promoter:50"
 /bound_moiety="BioB documented site"
 3411..3441
 /note="factor Sigma70: promoter bioB; documented +1 at
 808525"
 complement(3447..3473)
 /note="factor Sigma70: promoter bioA; documented +1 at
 808515"
 3493..4533
 /gene="b10B"
 /note="b0775"
 3493..4533
 /gene="b10B"
 /EC_number="2.8.1.-"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="o346: 99 pct identical to B10B_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPWYLSQVTELEFKPLDLLEBAQGVHQRHEDDPROVYST
 LLSIKTAGPCEDCKYCPQSSRYKTGLAEERLMEYQVLESARAKAGSTRFCGAM
 KNPHEDMPYLEOVNQGKAMGLEACWTGLTLESQORLANGLDYNNHNLPTSPER
 YGNIITTRVYOEIRIDLEKVRDAGIKVSGGIYGLGTPKVRDAGLLLOLANLPTPES
 VPIMLVKVGKTPPLADNDVDAPFRTIIVARIIMPTYSYVRLASAREONPETOAMC
 FMAGNSIFPGCKLITTPNEBEDKDLQFLKGLINPQTVAVLADNEDQORLEQALMT
 PDPEYNNMAL"
 4530..5684
 /gene="b10A"
 /note="b0776"
 4530..5684
 /gene="b10A"
 /EC_number="2.3.1.47"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="o384: 100 pct identical to B10F_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSMOEKINALDARRAADALRRRYVPAQAGRMVADDDROYLNF
 SSNDYGLSHHPQIIRAWOQGEQFISGSGSHVGVSVHQALEEFLEAEMVIGSRA
 LLFISGFAQNAVIAAMAKEDRIADRLSHASILEASTSPSOLRRFAHNDVTHLAR
 LLASPGQAOVYTEGVFSMDSDSAPLATIQQVTOOHNGMLVMDAHGTCVIGBOGRC
 SCWLQKVPPELLIVTFGKGGVSGAVALCSSIVADILLQFAHLITSTSPPAQOAL

Query Match 100.0%; Score 25; DB 1; Length 11022;
 Best Local Similarity 100.0%; Pred. No. 0.0018;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcaggtgaggtgacgacgcttg 25
 Db 3603 TCGTCAGGTGACGCTGACGACGCTTG 3627

RESULT 13
 SMABIO
 LOCUS 7215 bp DNA BCF 04-FEB-1999
 DEFINITION S. marcescens biotin operon, complete sequence.
 ACCESSION D17468
 VERSION D17468.1 GI:402530
 KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
 7-keto-8-amino-pelargonic acid synthetase; DAPA aminotransferase;
 DMB synthetase; KAPA synthetase; biotin operon; biotin synthetase;
 dehydrobiotin synthetase.
 Serratia marcescens (Strain:Str41) DNA.
 SOURCE
 ORGANISM
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;


```

REFERENCE      1 (bases 1 to 7215)
AUTHORS       Sakurai,N.
TITLE         Direct Submission
JOURNAL       Submitted (25-AUG-1993) to the DDBJ/EMBL/Genbank databases. Nooki
              Sakurai, Tanabe Selyaku Co., Ltd., Res Lab of Applied Biochemistry;
              2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan
              (E-mail:GFC01101@niftyserve.or.jp, nsakurai@dbj.nig.ac.jp,
              Tel:048-433-2545, Fax:048-433-2540)
REFERENCE      2 (bases 1 to 7215)
AUTHORS       Sakurai,N., Imai,Y., Akatsuka,H., Kawai,E., Komatsubara,S. and
              Tosa,T.
TITLE         Complete nucleotide sequence of biotin operon of Serratia
JOURNAL       marcescens
COMMENT       Unpublished (1993)
              Submitted (25-Aug-1993) to DDBJ by:
              Nooki Sakurai
              Res. Lab. of Applied Biochemistry
              Tanabe Selyaku Co., Ltd.
              2-50 Kawagishi-2-chome
              Toda, Saitama 335
              Japan
              Phone: 048-433-2545
              Email: nsakurai@dbj.nig.ac.jp
              Fax: 048-433-2540.

FEATURES
source        1..7215
              Location/Qualifiers
              /organism="Serratia marcescens"
              /strain="Sr41"
              /db_xref="taxon:615"
              complement(1353..2630)
              /gene="bioc"
              complement(1353..2630)
              /gene="biob"
              /EC-number="2.6.1.62"
              /codon_start=1
              /transl_table=11
              /product="7,8-diamino-pelargonic acid aminotransferase
              (DAPA aminotransferase)"
              /protein_id="BAA04284.1"
              /db_xref="GI:402531"
              /translation="MSVTSADLAFDQRIHIMPTSMKHPICYPIESAGVEQLADG
              RSLVDGSMMAAIIHGYNNPRLNOASQLEKMSHVMFGITHPAALISICRLIEMTP
              EALOCVPLADSGSAVEVSLKMALOYMOARERORILLRHGYHGDIFGAMVCDPD
              NSMSTLYOGYLAHPILFATAPQCRFDEWEEDDIPFALLHOGVEAAVILFEPVVG
              AGMRITHTYTKRRDYRRAIKILLIADDEIATGFGTGTGKPCFCAHOVVDILG
              KALTGVTLSATLTTRHVAETISGAAGCFMHGPTFNGFLACAVADASTALLAENR
              MQAOVSATETQLRELPLIGRCRMPHVRVIGAVGVMREPVDVAELDRGFVERGVA
              IRPEKLIYLMPPYIIIEAQLSRITAAVADAAR"
              2716..3756
              /gene="biob"
              2716..3756
              /gene="biob"
              /codon_start=1
              /transl_table=11
              /product="biotin synthetase"
              /protein_id="BAA04285.1"
              /db_xref="GI:402532"
              /translation="MAADRIHTVQAQALDEKPIELLFENQYTHROHPROYVS
              TLLSKTAGCPEDKCYCPQSPRYKTGLSESRIMOVQLESARKKAKAGSTRFCGAA
              MNPHEDMPYLQOVGVKAMGEMTCMLTGLDQTOAERLAELAGLDVYNNHLDTSPE
              FYGSTITTRSYERLDITLKVADAGIKVCSGIVLGGTVRRAGLIVQLNIPKPE
              SVPIMLVKGTGPLADNDVDPDFEIRITAVARIMMSSYRLSAGREONNEQTOAM
              CFMAGANSIFVGCCKLITTPNEDEKDLDFRLGLTNGPQTATHEHDNQCCVIAKOLL
              NADVAEFYNAAP"
              3756..4904
              /gene="biob"
              3756..4904
              /gene="biob"
              /codon_start=1
              /transl_table=11
              /product="7-keto-8-aminopelargonic acid synthetase (KAPA
              synthetase)"

REFERENCE      1 (bases 1 to 13501)
AUTHORS       Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postal,G., Hackel,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grothbeck,E.J., Davis,N.W., Lim,A., Dimmlanta,E., Potamoussis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
TITLE         Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

```

```

/protein_id="BAA04286.1"
/db_xref="GI:402533"
/translation="MSMOQRIFEOALAEERLNAAVRRROTEGSGROIRLGDRLXLPF
SGNDVILGISQDARVIAAMQOGRVGVGGSGHYTGSAHQAOLEBOLAMLGIPRL
LEITSYANQAVLAALAMQKGRILLADRLSHASILEAAGVAVARPAVPGQSTAGLARC
LAKPDGQRLAVTEBGLFSDMGDGAFLAEHLHRTTRAGAMLVADAHGIGVEGGRKS
CMQOGRPELIVATGKAFGVGAVALDEDAETAEILQFARHLIYSTAMPQAQCALQ
MALARIKSDDLARLONIRRFRCAGAPLALVTDSDPTAIQPLVCGNQRALDVR
LRECGIWSAIPRPVPPGARGRLRLTLTAHNSODIDRLLEVNDVSG"
4888..5655
/gene="bioc"
4888..5655
/gene="bioc"
/function="synthesis of plimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="the product of bioc"
/protein_id="BAA04287.1"
/db_xref="GI:402534"
/translation="MTSANDVYNKOAVASAFSRAAGSYDAALQDVGERRLGMSS
HGEQDLIDAGCGTGYESRMREGRKRVVALDAPGLVAVARORAAHHYLLGDIEVP
LPEAMDTCFSSLVQWMCSDLPALAEILRYTRPGCVILFSTLAGSLOELGDAMQOV
DERHYNALPLTQIRTAGCAVRLHELVELRLNPVYTLMLRSIKIGATHHNGRE
GGLMSRGRILAALQAAVPCRGQFPLSYHLAVGYRE"
5648..6331
/gene="biob"
5648..6331
/gene="biob"
/function="conversion from 7,8-diaminopelargonic acid into
dethiobiotin"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase (DTH synthetase)"
/protein_id="BAA04288.1"
/db_xref="GI:912453"
/translation="MSKRFPVGTCTDPEVKSVAASSALLQANRACYPSPAGKPVVAGS
EMTAEGLRNDALALQANSVALDYDEVNPFVFAETSPHIVSADGRITDARISDG
LRRLERADVILVAGAGGPTPLSAEYTDVWRQQLVILLVGIKLCINHAVALTA
QAVOAGLITLGIANDVGGAPGRHQEVLATLRMLPRCWAKSHTCRPERAPLQO
YLDLSILAC"

```

```

BASE COUNT    1347 a      2225 c      2367 g      1276 t
ORIGIN
Query Match      84.0%; Score 21; DB 1; Length 7215;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2      2      cgtcaagtgcagtcagtcagc 22
Db      2830  CGTCAGGTGCAGTCAGCAGC 2850

```

```

RESULT 14
AE005258      13501 bp      DNA      BCT      21-MAR-2001
LOCUS        Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
DEFINITION   of 155.
ACCESSION    AE005258 AE005174
VERSION      AE005258.1 GI:12513751
KEYWORDS
SOURCE
ORGANISM     Escherichia coli O157:H7 EDL933.
              Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postal,G., Hackel,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grothbeck,E.J., Davis,N.W., Lim,A., Dimmlanta,E., Potamoussis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
TITLE        Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

```

JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 13501)
AUTHORS Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,K.A. and Blattner,F.R.
TITLE Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Source location/Qualifiers
1..13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
<1..7576
/note="O-island #36; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933K; includes one copy of the 13 bp direct repeat that flanks the prophage"
66..665
/gene="lomK"
/note="20981"
66..665
/gene="lomK"
/function="putative membrane; other or unknown (Phage or prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from Genpept 118 : g117532789|gb|AAE63231.1|AF151091_2 (AF151091) lom [prophage P-Elba]"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage CP-933K"
/protein_id="AA65137.1"
/db_xref="GI:12513752"
/translation="MKVCAALLSAICLAAGVPMASHEOSTLSAGVLAHSTDPAGSDLLNINVKYREFTDILGITSFYANAEDDEKTHYSDTHMEDYVRNRFVSMAGPSVRVEMESAYAMAGVAVASRYSTSGDFRYTDKRRKTHVLITGSDARKSNISLAWGAGVQNPTESSAVVAVAYEXSGSGMDRTDGLVGYKPF"
719..2041
/gene="20982"
719..2041
/gene="20982"
/function="putative structure; structural component (phage or prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from Genpept 118 : g114585436|gb|AAD25464.1|AF125520.59 (AF125520) putative tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"
/protein_id="AA65138.1"
/db_xref="GI:12513753"
/translation="MGMAVQISGLKDGAKPIQNTIQLKARNSTXVVNTVASINPDGGRYSMEVYGGVSYTLVLEGFPSSHAGTISYEDSQPTLNDFLGAMTEDDARPEALRFQWVEEARHAEAKKNAAGEAETSAKNAGISAKSAEASAAADTSADASESARQAEASASAKSEESASSSSSAEAAQKASESIQATDAFLSKTAEASAGNARA TTSPEKRESASQASQASRIAAEDAVNRIPYVGPGRGKGECPGAPGPKDKKE RDTTPAGATGEGCGDTPGAPGAPGQCGEGTGTGTTNAGAPGQCGTGAAGPA GPGCKGTGAAGPYGATGPGPGGDEGTDIRFLGPMRLITNLSYGWEPGNDGAL I TGLTFLDPKADTQVQGMFQHLQVRFQGGPWQVKGDLDEVSQDGTGTE"
2004..2312
/gene="20984"
2004..2312
/gene="20984"
2004..2312
/gene="20984"
2004..2312
/gene="20984"

gene
CDS
2489..3469
/gene="20985"
2489..3469
/gene="20985"
/function="orf; other or unknown (Phage or Prophage Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from Genpept 118 : g116960367|gb|AAE33527.11 (AF170176) hypothetical protein predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AA65140.1"
/db_xref="GI:12513755"
/translation="MSPIRKTRHNSVNIQVSSPCQVTFAGKEVEIKYIDKPTPLFQWEPNPERKAKDEVPVTVNKRQHPYLDVNTNARIBSDNMGIFVGDGSSVQKTAFS KLEDFPNVMTIYREDVDFSMYDRKLSIDYHDIICEDRLTEDEKREYLLNLEKTEL EISKAQDSLISMVAKKRNHAMDFEFLNALAKAGEIFPCTYNTKNGISFGECGCTLD ADMILTKLGTIVAPDGISMNVDRNDSVNIENSALITVNSNHPALLEGISTPMHSVD AHEVYDGLGKGVAKYKFFFTPLNHNHCQDIEFNHRIILNITSQYTCSSW"
3530..4522
/gene="20986"
3530..4522
/gene="20986"
/function="orf; other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AA65141.1"
/db_xref="GI:12513756"
/translation="MKIPSIQSNFNSAPAGVYAPAPNRAVADYVLDIGKRLPLSAADLSVYVESVTRAVDRSRRLIDHYTDMIGNTYLDLAKSOTFRQAVSTGTHNEK VHTCCIKYRNEYELNESSYKTDIDTOSLTGNELEYDVQDELPFPCGAGENDNEPY VSFSAVDYDTSYEMPSMOWEGLHEIITHYTGSSDPGSDNIEGLPEILARVAOELG WSVPDFRGYAEPEEREAHLRLRLNALRQAMRHEENRFAFFRLGITSIRYBASPDFT EYSAVSNIIGYGFQOHDFFGLAINDMLDANOIQLYHGAPYIFTEGVDVKNHQO"
5350..6231
/gene="20989"
5350..6231
/gene="20989"
/function="orf; other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AA65142.1"
/db_xref="GI:12513757"
/translation="MISPYSVNLGCSWNSLTNLTSPDNRLVLSVDAVAHSDNGCAOV KYNQNRIRYVATDNKFCVTRESHSGCFNLHLRLGMPGEISRKTEJMLNNSPYSAAM ERQIVSNRPDLVPDYAPPELPSVDYVNLSPGVNICKGGAAYVDEADRTKYLKMF TTSQSENEVTSVCEFNQYVAGCAEKTGKNGGITIGIRMDKINGESLNTSSLPQO ABAHAYDMFRLREKGLIFVDTTETNVLITRAKNEFNPLDISSVNSDRSWSQIMQ STHGKQDILLIVLSKTI"
6462..7160
/gene="20990"

```

CDS
6462..7160
/gene="20990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG5143.1"
/db_xref="GI:12513758"
/translation="MPPSINILVHOSRSSMSDDIDSLVAKSSWV:KNNPOLRF
OGTDHNTYQOIEALDKIGSTFTGRVILNIAESIRKSTTVIHLNNSRLGYMAHD
IDAEHNGTSDHFHCLNVAEYPCGSGISVDFHATVHELLHFNNGELKAKES
SRAESOKYSPDLLEARTVGLAFSEEVISENKFHEEIGMPRTSTPXDALIHDDNT
VSJFOQVRLHAPL"
complement(7668..8144)
/gene="ybhb"
/note="20992"
complement(7668..8144)
CDS
complement(7668..8144)
/gene="ybhb"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAG5144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLKPRHFNHNGMGYDGNISPLAMDVDVPAITK
SFVYTCYDPAKPGSGMMHVVYVNLPAIDTVLPDGGSLVAMPDGVLOTPTGKIG
YDGAAPRKGERTHRYTFVHALDVRIVDEGCASAMGFVHFHSLASAITAMFS"
complement(8203..9492)
/gene="b10a"
/note="20993"
complement(8203..9492)
gene
CDS
complement(8203..9492)
Query Match
Best Local Similarity 100.0%; Score 21; DB 1; Length 13501;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 tcgtcaggtcagtcagtcac 21
|||||
Db 9689 TCGTCAGTGCAGTCAGTCAC 9709

RESULT 15
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
DEFINITION AP002553 BA000007
ACCESSION AP002553.1 GI:13360211
VERSION
KEYWORDS
SOURCE
. Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (sites)
REFERENCE
AUTHORS
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage vW1-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL
MEDLINE
20198780
2 (sites)
REFERENCE
AUTHORS
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an

```

```

JOURNAL
MEDLINE
20557356
3 (sites)
REFERENCE
AUTHORS
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage vW1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
JOURNAL
MEDLINE
20564182
4 (sites)
REFERENCE
AUTHORS
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
JOURNAL
MEDLINE
21156231
5 (bases 1 to 297816)
REFERENCE
AUTHORS
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:kengen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project
COMMENT
FEATURES
source
location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
/note="similar to SUCC_ECOLI g111786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="MNLHEOAKOLFARVGLPAPVGYACTPREAEASKIGAPVW
VKCOVHAGGKRGKGYVNVNSKEDIARFAENETGKRLVYQDANQOPNLLVEAT
DIAKELVLGAIVDRSSRRVFMASITBGEVEIKVABETPLHKKVALLDPLTGMPOG
RELAFLGLEKLVQOFTKLFMGILATLEERDLALETINPLVTKGDLICDGLGGA
DGNALFROPDLREMDOSQEDPREQAAGWELVYVADNGICGVNAGVAGLAWGTMDIV
KLGNLEPANFLVDVGATKERYTEAFKILISDQKRAVAVINIPGCIVRDLDAGTITG
AAVEGVNVVYVVRLEGNMELGAKKLADSGLNITIAKGLTDAQOQVAAVEGK"
1245..2114
/gene="ECs0754"
1245..2114
/gene="ECs0754"
/note="similar to SUCC_ECOLI g111786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSTILIDKNKRVICGQGTGSGTFRSQAATGATKMGVGTGPGK
GTTHGLGPFVNTVREAVAAATGATASVITVFAFPCKDSILEAIDAGIKLITITIGIPT

```

```

gene
CDS
    LDMITVAVKLDKAGVMTGNCPCGVITPGCKIIGIOPGHIHKPKWQIVSRSGTLTVE
    AVKOTIDYGGQSTCVGIGDPIGNSNFIDILEMFEKDPQTEAIVMIGETGSGAEEEA
    AAYIKETVPVYVYIAGVATPAGKRMHGAGAIAGGKGADEKFALEAAVAGKVTTRS
    LADIGALKITVLK*
    2781..3686
    /gene="ECS0755"
    2781..3686
    /note="probable transcriptional regulator, similar to
    transcriptional regulators e.g. glycine cleavage system
    transcription activator (gcv operon activator) -
    Escherichia coli g11417043|sp|p32064|GCV4_ECOLI percent
    identity 31 in 300 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative transcription regulator"
    /protein_id="BAB34178.1"
    /db_xref="GI:13360214"
    /translation="MGRKIPKTELLVTEVVARHESYTAABELATIGSAVFRQVSAL
    EERLRPLFSKSKRIPLNDTGKYYIGIVKEPLNKLEKDTNTMTWQPTVOVLEAVN
    PTFSTHMLIPNLEPFLKPDIVNIHSLANNGDPLNREDAVIMKRENCAMWAVEY
    LPEDEILPYOGSGLIAMSQKLSVAELTELPLEHSTRIITGMEEMFALSGVSPLEVN
    NGRPDLISMLIAAVRSNLGVALLPFAIQLHDLSDGWIIPCDVPIRTGNRIIMTWQF
    EKSDSPHLOOFREMLIAKSVQDEM"
    complement(3720..4322)
    /gene="ECS0756"
    complement(3720..4322)
    /note="probable cob(1)alamin adenosyltransferase, similar
    to cob(1)alamin adenosyltransferases (cordinoid
    adenosyltransferases) e.g. (Escherichia coli)
    g1115148|sp|p13040|BRUR_ECOLI percent identity 67 in 200
    aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative cob(1)alamin adenosyltransferase"
    /protein_id="BAB34179.1"
    /db_xref="GI:13360215"
    /translation="MEARISTEHWOROOKLEQVDTFVAAIEKKGILIVFTNGKGS
    KSTAAFGTVTRAVGKTVGVAQYIKGMDNCEVNIQLPGLVEPHIMGTGTEWQNR
    QADIDAAKSWSESKRMADKRYDLVDELTEYMLAYHLDTEEVYIASLQNRPAQOSV
    IYTGRCQSQILKMDPTVEIRPVKHAEDNGIQAGPGIDW"
    complement(4332..5984)
    /gene="ECS0757"
    complement(4332..5984)
    /note="probable fumarate hydratase, similar to fumarate
    hydratases e.g. fumarate hydratase class I, aerobic
    (fumarate) - Escherichia coli
    g11205981|sp|p00923|FUMA_ECOLI percent identity 68 in 545
    aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative fumarate hydratase"
    /protein_id="BAB34180.1"
    /db_xref="GI:13360216"
    /translation="MSKPTIOWPPIONKDKTEVTLISDQIITVELDGEVIRKIAPE
    AITLARQAFEASTFEFLSAHLOOVASILNDQOASSNKYVALDULNNAEVSAGVLP
    NCODITGTVIASKGQOITWIGNDNAEALSKGISTPENNLRFSQNAPLDWTYEVNTO
    TNLPAQIDISAVAGDEVHFLCVNKGSGSANKAAALYOEKSLILPEKLTAPLIIEKMSL
    GSAQCPYHIAFVVGISADQTLKIAKLASTKYVDNLPTSGNEGOGAFLPITIEKMSL
    EASQOFGIAGQFGKYEFAHDIRVIRLPLPHGSCPIIALMSASARINIKATINKKSTLU
    EKLEHNPQGVIPASIRENHQHOYDOLNRELRVMDLALPLPSTVYSISGLPIYAR
    DIAHAKIARIDSGEPMEYIKKHIIYTAGAKTPEEMWAGSLDPTTGGRRDGTIDTF
    QAAGCSLWLSKRNQOQVTDACHRHGGFNGSTGGAAALLAOEVYVSKRLCEVPELG
    MEAVMMMEVENLPAILVDDKGNMFESFOEDHRCASCPAGH"
    complement(6092..7372)
    /gene="ECS0758"
    complement(6092..7372)
    /gene="ECS0758"

```

```

/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. (Bacillus stearothermophilus)
g11121467|sp|p24943|bLTT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTMTLLALVLGMIIGVYLNNTASHPETAKLAQESITPT
TIFFLIKMIIAPLVVSTHLYVGIKMDAKALGRIFSKTLPLFICASLISALGILTY
NFMFGGGINFVAGCAETTGVALEPFLKIFSHAPTSIVAMANNETLOIYVSTI
FLGSLIAIGKSGSAIVAHADSLAHMLIKITGYMIFAPLITVRAISALAEGLAM
VSACIFMGEPEFTYMLLIVLLIGATLYVGCIRLRLKALISEPALLAFYSSSEAPF
GTLERLDFGVSPKIASVLPITGTSFNLVSGMSAPVAFIAQACNIHLISGEQITM
LILIMLTSKMGAVPRASMVVIAATLNQFNPEAGLILLMGVDFLDMGRSATVMSN
AMGAAVSRWGEHFGECRGKALKPNESVALP"
complement(7533..7850)
/gene="ECS0759"
complement(7533..7850)
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

Query Match 84.0%; Score 21; DB 1; Length 297816;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcgtcaggtgcaggtccacac 21
 DB 92421 TCGTACGGTGCAGGTCCACGAC 92441

```

RESULT 16
AE001862/C
LOCUS
DEFINITION
  AE001862 213732 bp DNA BCT 22-NOV-1999
  Deinococcus radiodurans R1 section 1 of 2 of the complete
  chromosome 2.
ACCESSION
  AE001862 AE001825
  AE001862.1 GI:6460468
VERSION
  AE001862.1
KEYWORDS
  SOURCE
  ORGANISM
    Deinococcus radiodurans.
    Deinococcus radiodurans.
    Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
REFERENCE
  1 (bases 1 to 213732)
  White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
  Dodson,R.J., Haft,D.H., Gwin,M.L., Nelson,W.C., Richardson,D.L.,
  Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
  Vamathevan,J.J., Lam,P., McDonald,L., Uitterback,T., Zaleski,C.,
  Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et al.
  Genome sequence of the radioreistant bacterium Deinococcus
  radiodurans R1
  Science 286 (5444), 1571-1577 (1999)
JOURNAL
  MEDLINE
  20036896
  2 (bases 1 to 213732)
  White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
  Dodson,R.J., Haft,D.H., Gwin,M.L., Nelson,W.C., Richardson,D.L.,
  Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
  Vamathevan,J.J., Lam,P., McDonald,L., Uitterback,T., Zaleski,C.,
  Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
  Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
  Smith,H.O., Venter,J.C. and Fraser,C.M.
  Direct Submission
  Submitted (02-NOV-1999) The Institute for Genomic Research, 9712
  Medical Center Dr, Rockville, MD 20850, USA
FEATURES
  source
    1..213732

```

```

gene
CDS
/organism="Deinococcus radiodurans"
/strain="R1"
/db_xref="taxon:1299"
/chromosome="2"
/gene="DRA0001"
653..1435
/feature="DRA0001"
/feature="similar to GB:D26185 SP:P37522 PID:467381
PID:580906 GB:AI009126 percent identity: 54.96; identified
by sequence similarity; putative"
/transl_table=1
/codon_start=1
/product="Chromosome partitioning ATPase, putative, Para
family"
/protein_id="AAFI2301.1"
/db_xref="GI:6460595"
/translation="MSAVKTLVFNHAGAGKTSLTPLNGVELAGRLVLLDLDP
OANTLWIGTIGSTREMTYVPAVDGOLPSPKAPGLDVI PAHVSLAIVEGOMGV
GAGRLRRALAEVSGDVALIDSPSLGOLAIILAALAOIMTYVPYTRKGLDALPG
LQGLTEREVRDPLVALYVTFYDARRRHDQEVLDLKAHLSPLARVPQREAVWL
DSTAQAPVSEYAPGTPVHADVQRLTADIAALGVAVGENA"
1432..2313
/gene="DRA0002"
1432..2313
/feature="DRA0002"
/feature="similar to GB:L34077 GB:X85964 PID:407375
PID:757761 PID:1163135 percent identity: 57.34; identified
by sequence similarity; putative"
/transl_table=1
/codon_start=1
/product="Chromosome partitioning protein, ParB family"
/protein_id="AAFI2300.1"
/db_xref="GI:6460594"
/translation="MTRRRPERRDLCLLGETPVDSQANDIRALPVNELKVGSTP
RRSEDLERLESLAESTIRAHGVLPPLVSVGVQYEVAVGERMRRAQALAEVAVV
RSLNEQARAALLENLORNI NVIDEVSGLELIALTLGLEREARRLAQILRAV
GDEHQLDOVFRSGEMTFRFAKRLRLMNPOLPRAGLPLTIGSVASAPER
QALLKLQNGASRQLQALQTSQTSANVPEHFAYKLSKRLSLDPTREALDR
WLARPERVQALDEQS"
complement(2418..3881)
/gene="DRA0003"
complement(2418..3881)
/feature="DRA0003"
/feature="similar to GB:M88334 SP:P25526 PID:147901 GB:U00096
PID:1789015 percent identity: 69.60; identified by
sequence similarity; putative"
/transl_table=1
/codon_start=1
/product="succinate-semialdehyde dehydrogenase"
/protein_id="AAFI2294.1"
/db_xref="GI:6460588"
/translation="MTTDLRTVYSSVTRSQAYFDGEWRNA PRNFVBRPGNEVIGE
VADCTPTDARQALDAEVALREMOVNEYERKILRMHDLMEHKEBELADLMTLEMG
KPISETRGVHVAASETIEWCAEAGRIAGERINTLRFPRKGLTISEPVGLVYAVPMN
PRAGITTKRAAPALACGVMILKPAELSPALTYLTPLMLKAGCPANFOVLPYNDA5
ALTOPPANDSRKLTFTGSTEVEGLILVQOAGTIKKVSLGLGHAPLVDDALER
AASEVQSKFRNSGTCVCTNRVYVORGVAEFTLLEKTRADLDGSPFDEATQCP
VVEQGLDKVQOVODALITKAQATTTGGVSSGLFPQPLVLDVAPDSLLRETEGP
VAPVITFTEEEGLRLANDSEYGLAAVATRDLAGFRIAEGLIEGIVGINDGLPSSA
APHVPGGKMKNSGVREGHMLVEYLLETKFVSLGS"
4024..4668
/gene="DRA0004"
4024..4668
/feature="DRA0004"
/feature="similar to GP:2924378 percent identity: 55.29;
identified by sequence similarity; putative"
/transl_table=1
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="AAFI2299.1"
/db_xref="GI:6460593"
/translation="MNSLSLAVALIMAGSTRMGQPKALPLPGMPLCRHIAAQTIVE

```

```

gene
CDS
AGCGEYAVLPPGETGDAVRALQGLPIRWVYNPOPAQGLSLSFQAGALWABESAP
AVPPTLADPLVRDTHOALGTFCGAEQPAATITFRYGVAPPTLRLPRLEPP
PSDHCPTLRLRELGREIKIDRRABELLIDIPRMAQAQQLFVEVTPGAGDDP"
complement(4719..5897)
/gene="DRA0005"
complement(4719..5897)
/feature="DRA0005"
/feature="similar to GB:L33464 SP:P47734 PID:496118 percent
identity: 74.23; identified by sequence similarity;
putative"
/transl_table=1
/codon_start=1
/product="alcohol dehydrogenase, zinc-containing"
/protein_id="AAFI2293.1"
/db_xref="GI:6460587"
/translation="KKAIVQGVKNKGVETVPDPLLPDPAIVKITSATGSDPLH
LDGYIPSMETGDLGHEVEVEVSEVRKIKVDRVYVPPNINAGCDPCKRRFES
ACDNSPNHRMAEALYVAGSGALFVSHMYGVAGAQVAVRVPADVPKLTESGL
KDEOVFLDIFPTGYQAAEQCNIVPGRVAVFAGAPGVQFARSQMLGAHVI
DRVERILAMAEAGCOTTINERKEDVLLSREATGGRPDHVIDAVGMEAHGPGSTM
DKVOOKLITFPRITLALMAIILSCAKGTVSMGVYGGVLDVPIGAAPAKGLIRMG
QTHTHRIITSLGRITAGQIDPSFTVTHNASIDQAPELKTFPRKQDCCIKVILPWA
"
complement(5894..6700)
/gene="DRA0006"
complement(5894..6700)
/feature="DRA0006"
/feature="similar to PID:1653488 percent identity: 48.46;
identified by sequence similarity; putative"
/transl_table=1
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="AAFI2292.1"
/db_xref="GI:6460586"
/translation="MDRRSSPOLRTYTNMNRQAPAPACATGNSPGATMQLRSL
TDVAVGALAAAGLRSSGLKVVLTGIGKGLIVARCONPLATKIDQNDGEVY
RGAVYTGKPRVAVSWWRDLPGLPLMTHLRSVEYLDLDRSRMTAPAPLDTGSEA
ELTNADESRQILAMRSLEPGARIENTSGYLFEPAPGARGTEVVALTYRPGGSAVIA
RMNDEFSQQLRDLMLFRKREQELGHAPTAQOGSSRAALTPPNKRGDLY"
6839..7744
/gene="DRA0007"
6839..7744
/feature="DRA0007"
/feature="similar to GB:AI009126 percent identity: 62.98;
identified by sequence similarity; putative"
/transl_table=1
/codon_start=1
/product="ABC transporter, ATP-binding protein, putative"
/protein_id="AAFI2291.1"
/db_xref="GI:6460585"
/translation="MITTFQVSKTYGHVATLSDPNLTLTGELTALFPNGKSKSTA
IGILLGLSAPSGVRYLGADPPRRDVARIGAMPQESALPALTVEAVTLFASFP
APLCTVDALALADLGPVAGRRARQLSSGQKRRALFALAVYGDPELLIDETPTGMAO
SRAFWAVAVGLRGRGRTLLTTHYIEAKRTDRVYVMNGKILADDTPOGIRSSVG
GARVSEVSDIQAQLELRLPGVSAQVDAAGRADLRKTSVPALLALIGSTTSDLEV
RRALIEAYVQLTGPQDMTAVTRSA"
7861..8652
/gene="DRA0008"
7861..8652
/feature="DRA0008"
/feature="similar to GB:AI009126 percent identity: 52.46;
identified by sequence similarity; putative"
/transl_table=1
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="AAFI2290.1"
/db_xref="GI:6460584"
/translation="MTTLPNVNPARRTSALPVLAQLTLGELRRLLRSPMTVGAITGPV
MFAFGLPAVOEGATDPHGVILTQFAVSLSLALFSFGAAVATERSSGMRLRL
RSSPLPVLYVSKTIALAAGAVSLTALFAHFAGAVTLPLDALALLAGKLLGLMI
PLVATGICIGLASPOAAOILANLISVMSASGLFVPLDLPREFVOIAPLLPAYIV
SOLATNVYSGTASEPAMHIALAATFIVGFLAVWGLKRESRSG"
8609..9819

```


[illegible]

JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

2 (bases 1 to 69900)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Abramson, H., Allen, N., Boguslavskiy, I., Bonkhalter, B., Brown, A., Burkett, G., Castle, A., Chepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Deatrel, A., K., Dewar, K., Domino, M., Doyle, M., Fenesior, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardys, S., Grant, G., Haags, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Liu, C., Lu, G., Locke, K., Macdonald, P., Marguis, N., Mcwan, P., McGurk, A., McKernan, K., McPeckers, R., Meldrum, J., Meneses, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K., Pierre, N., Plisani, C., Pollard, P., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testifye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., & Zody, M.

Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced g1:67121397.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L3201
Center clone name: 47.D.19

* NOTE: This record contains 79 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1	787:	contig of 787 bp	in length
*	888	gap of 100 bp	
*	888	1651:	contig of 764 bp in length
*	1552	1751:	gap of 100 bp
*	1752	2546:	contig of 795 bp in length
*	2547	2646:	gap of 100 bp
*	2649	3438:	contig of 792 bp in length
*	3439	3538:	gap of 100 bp
*	3539	4327:	contig of 789 bp in length
*	4328	4427:	gap of 100 bp
*	4428	5234:	contig of 807 bp in length
*	5235	5334:	gap of 100 bp
*	5335	6129:	contig of 795 bp in length
*	6130	6229:	gap of 100 bp
*	6230	7040:	contig of 811 bp in length
*	7041	7140:	gap of 100 bp
*	7141	7931:	contig of 791 bp in length
*	7932	8031:	gap of 100 bp
*	8032	8810:	contig of 779 bp in length
*	8811	8910:	gap of 100 bp
*	8911	9696:	contig of 786 bp in length
*	9697	9796:	gap of 100 bp
*	9797	10586:	contig of 800 bp in length
*	10597	10696:	gap of 100 bp
*	10697	11493:	contig of 797 bp in length
*	11494	11593:	gap of 100 bp
*	11594	12399:	contig of 806 bp in length

```

* 12400 12499: gap of 100 bp
* 12500 13296: contig of 797 bp in length
* 13297 13396: gap of 100 bp
* 13397 14152: contig of 756 bp in length
* 14153 14252: gap of 100 bp
* 14253 15034: contig of 782 bp in length
* 15035 15134: gap of 100 bp
* 15135 15922: contig of 788 bp in length
* 15923 16022: gap of 100 bp
* 16023 16795: contig of 773 bp in length
* 16796 16895: gap of 100 bp
* 16896 17698: contig of 803 bp in length
* 17699 17798: gap of 100 bp
* 17799 18568: contig of 770 bp in length
* 18569 18668: gap of 100 bp
* 18669 19457: contig of 789 bp in length
* 19458 19557: gap of 100 bp
* 19558 20342: contig of 785 bp in length
* 20343 20442: gap of 100 bp
* 20443 21239: contig of 797 bp in length
* 21240 21339: gap of 100 bp
* 21340 22122: contig of 783 bp in length
* 22123 22222: gap of 100 bp
* 22223 22995: contig of 773 bp in length
* 22996 23095: gap of 100 bp
* 23096 23836: contig of 741 bp in length
* 23837 23936: gap of 100 bp
* 23937 24720: contig of 784 bp in length
* 24721 24820: gap of 100 bp
* 24821 25590: contig of 770 bp in length
* 25591 25690: gap of 100 bp
* 25691 26480: contig of 790 bp in length
* 26481 26580: gap of 100 bp
* 26581 27368: contig of 788 bp in length
* 27369 27468: gap of 100 bp
* 27469 28258: contig of 790 bp in length
* 28259 28358: gap of 100 bp
* 28359 29134: contig of 776 bp in length
* 29135 29234: gap of 100 bp
* 29235 30016: contig of 782 bp in length
* 30017 30116: gap of 100 bp
* 30117 30897: contig of 781 bp in length
* 30898 30997: gap of 100 bp
* 30998 31770: contig of 773 bp in length
* 31771 31870: gap of 100 bp
* 31871 32662: contig of 792 bp in length
* 32663 32762: gap of 100 bp
* 32763 33558: contig of 796 bp in length
* 33559 33658: gap of 100 bp
* 33659 34448: contig of 790 bp in length
* 34449 34548: gap of 100 bp
* 34549 35346: contig of 798 bp in length
* 35347 35446: gap of 100 bp
* 35447 36251: contig of 805 bp in length
* 36252 36351: gap of 100 bp
* 36352 37136: contig of 785 bp in length
* 37137 37236: gap of 100 bp
* 37237 38021: contig of 785 bp in length
* 38022 38121: gap of 100 bp
* 38122 38908: contig of 787 bp in length
* 38909 39008: gap of 100 bp
* 39009 39800: contig of 792 bp in length
* 39801 39900: gap of 100 bp
* 39901 40683: contig of 783 bp in length
* 40684 40783: gap of 100 bp
* 40784 41551: contig of 766 bp in length
* 41552 41651: gap of 100 bp
* 41652 42409: contig of 736 bp in length
* 42410 42509: gap of 100 bp
* 42510 43302: contig of 793 bp in length
* 43303 43402: gap of 100 bp
* 43403 44204: contig of 802 bp in length
* 44205 44304: gap of 100 bp

```

```

Query Match      68.0% Score 17: DB 2: Length 69900;
Best Local Similarity 100.0%; Pred. No. 22.
Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2  cytcaggtcaggtcag 18
Db      19825  COTCAGTGCAGTGCAG 19809

```

```

RESULT 20
AC069132

```

```

LOCUS      AC069132 151340 bp DNA HTG 17-JUN-2000
DEFINITION Homo sapiens chromosome 5 clone RP11-55618, *** SEQUENCING IN
PROGRESS ***. 62 unordered pieces.

```

```

ACCESSION AC069132.2 GI:8469032
VERSION    AC069132.2
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.

```

```

ORGANISM   Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 151340)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
            JOURNAL
            2 (bases 1 to 151340)
            Waterston,R.H.
            REFERENCE
            TITLE
            JOURNAL

```

```

Submitted (18-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```


COMMENT On Jun 12, 2000 this sequence version replaced gi:7924001.

----- Genome Center -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1502: contig of 1502 bp in length
1503 1602: gap of unknown length
1603 2741: contig of 1139 bp in length
2742 2841: gap of unknown length
2842 4036: contig of 1195 bp in length
4037 4136: gap of unknown length
4137 5620: contig of 1484 bp in length
5621 5720: gap of unknown length
5721 7080: contig of 1360 bp in length
7081 7180: gap of unknown length
7181 8839: contig of 1659 bp in length
8840 8939: gap of unknown length
8940 9987: contig of 1048 bp in length
9988 10088: gap of unknown length
10089 11156: contig of 1069 bp in length
11157 11256: gap of unknown length
11257 12404: contig of 1148 bp in length
12405 12504: gap of unknown length
12505 13861: contig of 1357 bp in length
13862 13961: gap of unknown length
13963 15241: contig of 1280 bp in length
15242 15341: gap of unknown length
15342 17009: contig of 1668 bp in length
17010 17109: gap of unknown length
17110 18897: contig of 1788 bp in length
18898 18997: gap of unknown length
18999 20389: contig of 1392 bp in length
20390 20489: gap of unknown length
20490 22042: contig of 1553 bp in length
22043 22142: gap of unknown length
22143 24057: contig of 1915 bp in length
24058 24157: gap of unknown length
24158 26189: contig of 2032 bp in length
26190 26289: gap of unknown length
26290 27959: contig of 1670 bp in length
27960 28059: gap of unknown length
28060 30067: contig of 2008 bp in length
30068 30167: gap of unknown length
30168 32456: contig of 2289 bp in length
32457 32556: gap of unknown length
32557 34597: contig of 2041 bp in length
34598 34697: gap of unknown length
34698 36611: contig of 1914 bp in length
36612 36711: gap of unknown length
36712 38677: contig of 1966 bp in length
38678 38777: gap of unknown length
38779 40974: contig of 2197 bp in length
40975 41074: gap of unknown length
41075 42666: contig of 1592 bp in length
42667 42766: gap of unknown length
42767 44292: contig of 1526 bp in length
44293 44392: gap of unknown length
44393 46032: contig of 1640 bp in length
46033 46132: gap of unknown length
46133 48815: contig of 2663 bp in length
48816 48915: gap of unknown length
48916 51058: contig of 2143 bp in length
51059 51158: gap of unknown length
51159 53094: contig of 1936 bp in length
53095 53194: gap of unknown length
53195 54724: contig of 1530 bp in length
54725 54824: gap of unknown length

54825 57019: contig of 2195 bp in length
57020 57119: gap of unknown length
57120 59736: contig of 2617 bp in length
59737 59836: gap of unknown length
59837 61510: contig of 1674 bp in length
61511 61610: gap of unknown length
61611 63247: contig of 1637 bp in length
63248 63347: gap of unknown length
63348 65262: contig of 1915 bp in length
65263 65362: gap of unknown length
65363 67493: contig of 2131 bp in length
67494 67593: gap of unknown length
67594 70740: contig of 3147 bp in length
70741 70840: gap of unknown length
70841 72256: contig of 1416 bp in length
72257 72356: gap of unknown length
72357 73456: contig of 3100 bp in length
73457 75556: gap of unknown length
75557 77328: contig of 1772 bp in length
77329 77428: gap of unknown length
77429 79851: contig of 2423 bp in length
79852 79951: gap of unknown length
79952 82877: contig of 2926 bp in length
82878 82977: gap of unknown length
82978 84859: contig of 1882 bp in length
84860 84959: gap of unknown length
84960 87856: contig of 2897 bp in length
87857 87956: gap of unknown length
87957 91677: contig of 3721 bp in length
91678 91777: gap of unknown length
91779 94536: contig of 2759 bp in length
94537 94636: gap of unknown length
94637 97377: contig of 2741 bp in length
97378 97477: gap of unknown length
97478 100981: contig of 3504 bp in length
100982 101081: gap of unknown length
101082 103958: contig of 2877 bp in length
103959 104058: gap of unknown length
104059 106903: contig of 2845 bp in length
106904 107003: gap of unknown length
107004 109662: contig of 2659 bp in length
109663 109762: gap of unknown length
109763 113672: contig of 3910 bp in length
113673 113772: gap of unknown length
113773 118353: contig of 4581 bp in length
118354 118453: gap of unknown length
118454 121541: contig of 3088 bp in length
121542 121641: gap of unknown length
121642 124122: contig of 2481 bp in length
124123 124222: gap of unknown length
124223 128114: contig of 3892 bp in length
128115 128214: gap of unknown length
128215 131684: contig of 3470 bp in length
131685 131784: gap of unknown length
131785 135575: contig of 3791 bp in length
135576 135675: gap of unknown length
135676 140587: contig of 4912 bp in length
140588 140687: gap of unknown length
140688 145142: contig of 4455 bp in length
145143 145242: gap of unknown length
145243 151340: contig of 6098 bp in length.

FEATURES
source 1. 151340
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-55618"

BASE COUNT 43058 a 29029 c 29780 g 43356 t 6117 others
ORIGIN

Query Match 68.0%; Score 17; DB 2: Length 151340;
Best Local Similarity 100.0%; Pred. No. 19;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 cgtcaggtcaggtcag 18
 |||
 Db 93019 CGTCAGGTGCGAGTCAG 93035

RESULT 21
 AC068100/c 161742 bp DNA HTG 04-JUL-2001
 LOCUS Homo sapiens chromosome X clone RP11-812N12 map X, WORKING DRAFT
 DEFINITION
 AC068100
 ACCESSION
 AC068100.5 GI:14595922
 VERSION
 KEYWORDS
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 1 (bases 1 to 161742)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Unpublished
 2 (bases 1 to 161742)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Unpublished

Anderson, S., Baldwin, J., Barne, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Bouknight, B., Brown, A., Burkelt, G.,
 Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Lakoque, K., Lamazares, R., Landers, T., Lehotzky, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
 Meldrum, J., Menus, L., Mihova, T., Miranda, C., Miura, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Poltara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Vassiliou, S., Theodorou, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainou, J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (28-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 4, 2001 this sequence version replaced g1:13357469.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu

Project Information
 Center project name: L7131
 Center clone name: 812_N_12

Summary Statistics

Sequencing vector: M13; M77815; 4% of reads
 Sequencing vector: Plasmid; n/a; 96% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731

Consensus quality: 158467 bases at least Q40
 Consensus quality: 159785 bases at least Q30
 Consensus quality: 160319 bases at least Q20
 Insert size: 165000; agarose-fp
 Insert size: 160742; sum-of-coverage

Quality coverage: 6.9 in Q20 bases; agarose-fp
 Quality coverage: 7.1 in Q20 ba.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1586: contig of 1586 bp in length
 * 1587 1686: gap of 100 bp
 * 1687 4904: contig of 3218 bp in length
 * 4905 5004: gap of 100 bp
 * 5005 6284: gap of 1280 bp in length
 * 6285 6384: gap of 100 bp
 * 6385 87528: contig of 81144 bp in length
 * 87529 87628: gap of 100 bp
 * 87629 90089: contig of 2461 bp in length
 * 90090 90189: gap of 100 bp
 * 90190 96930: contig of 6741 bp in length
 * 96931 97030: gap of 100 bp
 * 97031 103279: contig of 6249 bp in length
 * 103280 103379: gap of 100 bp
 * 103380 114728: contig of 11349 bp in length
 * 114729 114828: gap of 100 bp
 * 114829 125622: contig of 10794 bp in length
 * 125623 125722: gap of 100 bp
 * 125723 156844: contig of 31122 bp in length
 * 156845 156944: gap of 100 bp
 * 156945 161742: contig of 4798 bp in length.

FEATURES
 SOURCE
 1. 161742
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="X"
 /clone="RP11-812N12"
 /clone_lib="RCR1-11 Human Male BAC"
 1. 1586
 /note="assembly-fragment
 clone_end:SP6
 vector_side:left"
 1687. 4904
 /note="assembly-fragment"
 5005. 6284
 /note="assembly-fragment"
 6385. 87528
 /note="assembly-fragment"
 87629. 90089
 /note="assembly-fragment"
 90190. 96930
 /note="assembly-fragment"
 97031. 103279
 /note="assembly-fragment"
 103380. 114728
 /note="assembly-fragment"
 114829. 125622
 /note="assembly-fragment"
 125723. 156844
 /note="assembly-fragment"
 156945. 161742
 /note="assembly-fragment
 clone_end:r7
 vector_side:right"

misc_feature
 /note="assembly-fragment
 clone_end:SP6
 vector_side:left"
 1687. 4904
 /note="assembly-fragment"
 5005. 6284
 /note="assembly-fragment"
 6385. 87528
 /note="assembly-fragment"
 87629. 90089
 /note="assembly-fragment"
 90190. 96930
 /note="assembly-fragment"
 97031. 103279
 /note="assembly-fragment"
 103380. 114728
 /note="assembly-fragment"
 114829. 125622
 /note="assembly-fragment"
 125723. 156844
 /note="assembly-fragment"
 156945. 161742
 /note="assembly-fragment
 clone_end:r7
 vector_side:right"
 BASE COUNT 45576 a 33084 c 34118 g 47963 t 1001 others
 ORIGIN

Query Match 68.0%; Score 17; DB 2; Length 161742;
 Best local Similarity 100.0%; Pred No. 19;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 cgtcaggtcaggtcag 18
 |||
 Db 70155 CGTCAGGTGCGAGTCAG 70139

RESULT 22
 AC041048/c
 LOCUS
 DEFINITION
 AC041048 168360 bp DNA HTG 07-JUL-2000
 SEQUENCE. 17 unordered pieces.
 AC041048.3 GI:7770701
 VERSION HTG: HTGS_PHASE1: HTGS_DRAFT.
 KEYWORDS
 SOURCE human.
 ORGANISM
 Eukaryote: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
 1 (bases 1 to 168360)
 Waterston, R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 168360)
 Waterston, R.H.
 Direct Submission
 Submitted (11-APR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On May 12, 2000 this sequence version replaced g1:752530.

Genome Center

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Project Information

Center project name: H.NH0759J14

Summary Statistics

Sequencing vector: M13; 100%
 Sequencing vector: plasmid; 0%
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 157456 bases at least Q40
 Consensus quality: 163068 bases at least Q20
 Insert size: 173000; agarose-fp
 Insert size: 166760; sum-of-contigs
 Quality coverage: 4.08 in Q20 bases; agarose-fp
 Quality coverage: 4.27 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1711: contig of 1711 bp in length
 * 1712 1811: gap of unknown length
 * 1812 6135: contig of 4324 bp in length
 * 6136 6236: gap of unknown length
 * 6236 8687: contig of 2452 bp in length
 * 8688 8787: gap of unknown length
 * 8788 11647: contig of 2860 bp in length
 * 11648 11747: gap of unknown length
 * 11748 13718: contig of 1971 bp in length
 * 13719 13818: gap of unknown length
 * 13819 18016: gap of 4158 bp in length
 * 18017 18116: gap of unknown length
 * 18117 21570: contig of 3454 bp in length
 * 21571 21670: gap of unknown length
 * 21671 24723: contig of 3053 bp in length
 * 24724 24823: gap of unknown length
 * 24824 32650: contig of 7827 bp in length
 * 32651 32750: gap of unknown length
 * 32751 46491: contig of 13741 bp in length

* 46492 46591: gap of unknown length
 * 46592 60127: contig of 13536 bp in length
 * 60128 60227: gap of unknown length
 * 60228 74481: contig of 14254 bp in length
 * 74482 74582: gap of unknown length
 * 74583 90887: contig of 16305 bp in length
 * 90888 104984: gap of unknown length
 * 104985 127379: gap of unknown length
 * 127380 127479: gap of 22295 bp in length
 * 127480 147028: contig of 19549 bp in length
 * 147029 147128: gap of unknown length
 * 147129 168360: contig of 21232 bp in length.

FEATURES

Source
 1. 168360
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-759J14"
 1. 1711
 /note="assembly_name:Contig6"
 1812. 6135
 /note="assembly_name:Contig7"
 6236. 8687
 /note="assembly_name:Contig8
 clone_end:SP6
 vector_side:right"
 8788. 11647
 /note="assembly_name:Contig9"
 11748. 13718
 /note="assembly_name:Contig10"
 13819. 18016
 /note="assembly_name:Contig11"
 18117. 21570
 /note="assembly_name:Contig12"
 21671. 24723
 /note="assembly_name:Contig13"
 24824. 32650
 /note="assembly_name:Contig14"
 32751. 46491
 /note="assembly_name:Contig15"
 46592. 60127
 /note="assembly_name:Contig16"
 60228. 74481
 /note="assembly_name:Contig17"
 74582. 90886
 /note="assembly_name:Contig18"
 90987. 104984
 /note="assembly_name:Contig19"
 105085. 127379
 /note="assembly_name:Contig20"
 127480. 147028
 /note="assembly_name:Contig21"
 147129. 168360
 /note="assembly_name:Contig22"
 misc_feature
 BASE COUNT 48985 a 34329 c 34643 g 48785 t 1618 others
 ORIGIN

Query Match 68.0% Score 17: DB 2: Length 168360;
 Best Local Similarity 100.0% Pred. No. 19;
 Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cctcaggtcaggtcag 18
 |||
 Db 153451 CGTCAGGTGAGGTCAG 153435

RESULT 23
 AC060833/c
 LOCUS
 DEFINITION Homo sapiens chromosome 4 clone RP11-556F3, WORKING DRAFT SEQUENCE,

4 unordered pieces.
 AC060833.3 GI:9838272
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 176379)
 Waterston, R.H.
 The sequence of Homo sapiens clone
 unpublished
 2 (bases 1 to 176379)
 Waterston, R.H.
 Direct Submission
 Submitted (20-APR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Aug 17, 2000 this sequence version replaced gi:8980022.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H.NH0556F03
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Chemistry: Dye-terminator; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 173730 bases at least Q40
 Consensus quality: 175046 bases at least Q30
 Insert size: 174000; agarose-fp
 Insert size: 177415; sum-of-ctnigs
 Quality coverage: 6.66 in Q20 bases; sum-of-ctnigs
 Quality coverage: 6.55 in Q20 bases; sum-of-ctnigs
 ----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1419: contig of 1419 bp in length
 * 1420 1519: gap of unknown length
 * 1520 10016: contig of 8497 bp in length
 * 10017 10116: gap of unknown length
 * 10117 79219: contig of 69103 bp in length
 * 79220 79319: gap of unknown length
 * 79320 176379: contig of 97060 bp in length.
 Location/Qualifiers
 1. 176379
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /clone="RP11-556F3"
 misc_feature
 1. 1419
 /note="assembly_name:Contig5"
 misc_feature
 1520. 10016
 /note="assembly_name:Contig6
 clone_end:sp6
 vector_side:right"
 10117. 79219
 /note="assembly_name:Contig7"
 79320. 176379
 /note="assembly_name:Contig8
 clone_end:t7
 vector_side:left"

BASE COUNT 49569 a 33984 c 35497 g 57025 t 304 others
 ORIGIN
 Query Match 68.0%; Score 17; DB 2; Length 176379;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 cgtcagtgacaggtcag 18
 |||
 Db 81954 CGTCAGTGCAGGTCAG 81938

RESULT 24
 AC022781/c
 LOCUS
 DEFINITION
 AC022781.6 GI:14595917
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
 AC022781
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
 house mouse.
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 220166)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Aldrich, E.,
 Anderson, S., Baldwin, J., Barlow, N., Beckwith, R., Beda, F.,
 Boguski, M.S., Brown, A., Burkett, G., Castelle, A.,
 Choquet, Y., Collins, S., Collins, S., Collins, S., Cooke, P.,
 DeRubeis, R., Dewar, K., Domino, M., Doyle, M., Fennell, J.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Hagos, B., Heald, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
 Landers, T., Lechoczky, J., Levine, R., Lien, G., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGuck, A., McKernan, K.,
 McPeckers, R., Meldrum, J., Meneses, L., Morrow, J., Naylor, J.,
 Norman, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Telford, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 4, 2001 this sequence version replaced gi:10305259.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- genome center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information -----
 Center project name: L5913
 Center clone name: 199.H.17
 ----- Summary Statistics -----
 Sequencing vector: M13; M7815; 43% of reads
 Sequencing vector: Plasmid; n/a; 57% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 217289 bases at least Q40
 Consensus quality: 218800 bases at least Q30
 Insert size: 206000; agarose-fp
 Quality coverage: 9.5 in Q20 bases; agarose-fp

TITLE
 JOURNAL
 COMMENT

Quality coverage: 8.9 in Q20 b.

* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 22567: contig of 22567 bp in length
* 22568 22667: gap of 100 bp
* 22668 28303: contig of 5636 bp in length
* 28304 28403: gap of 100 bp
* 28404 109561: contig of 81358 bp in length
* 109562 109661: gap of 100 bp
* 109662 118270: contig of 8609 bp in length
* 118271 118370: gap of 100 bp
* 118371 140406: contig of 22036 bp in length
* 140407 140506: gap of 100 bp
* 140507 158389: contig of 17883 bp in length
* 158390 158489: gap of 100 bp
* 158490 211583: contig of 53094 bp in length
* 211584 211683: gap of 100 bp
* 211684 220166: contig of 8483 bp in length.
```

FEATURES

SOURCE

```
1. 220166
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /chromosome="11"
  /map="11"
  /clone="RP23-199H17"
  /clone_lib="RPC1-23 Female Mouse BAC"
```

misc_feature

```
1. 22567
  /note="assembly_fragment"
  /clone_end:SP6
  vector_side:left"
```

misc_feature

```
22668. 28303
  /note="assembly_fragment"
  28404. 109561
  /note="assembly_fragment"
```

misc_feature

```
109662. 118270
  /note="assembly_fragment"
  118371. 140406
  /note="assembly_fragment"
```

misc_feature

```
140507. 158389
  /note="assembly_fragment"
  158490. 211583
  /note="assembly_fragment"
```

misc_feature

```
211684. 220166
  /note="assembly_fragment"
  /clone_end:r7
  vector_side:right"
```

misc_feature

```
BASE COUNT 53983 a 54774 c 55714 g 54982 t 713 others
ORIGIN
```

Query Match

Best Local Similarity 68.0%; Score 17; DB 2; Length 220166;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 cgtcagatcagatcag 18
|||||

Db 37316 CGTCAGTCAGTCAG 37300

RESULT 25

AL591065

LOCUS 222212 bp DNA HTG 13-JUN-2001
DEFINITION Mus musculus chromosome 11 clone RP23-199H17, *** SEQUENCING IN

ACCESSION

AL591065.2 GI:13990285

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 222212)
Plumb,B.
Direct Submission
Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On May 7, 2001 this sequence version replaced gi:13938846.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: bml99H17
----- Summary Statistics
Assembly program: XCAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 218929 bases at least Q40
Consensus quality: 219403 bases at least Q30
Consensus quality: 219831 bases at least Q20
Insert size: 220512; sum-of-contigs
Insert size: 221854; 2.3% error; agarose-fp
Quality coverage: 10.86x in Q20 bases; sum-of-contigs Quality
coverage: 11.12x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

SOURCE

```
1. 222212
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /chromosome="11"
  /clone="RP23-199H17"
  /clone_lib="RPC1-23"
  1. 19515
  /note="assembly_fragment:03069
  fragment_chain:1"
```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature 119149..139095
/note="assembly_fragment:04963
fragment_chain:1"
misc_feature 139196..141828
/note="assembly_fragment:02943
fragment_chain:1"
misc_feature 141929..150289
/note="assembly_fragment:04561
fragment_chain:2"
misc_feature 150390..166695
/note="assembly_fragment:06222
fragment_chain:2"
misc_feature 166796..169310
/note="assembly_fragment:04338
fragment_chain:2"
misc_feature 169411..210287
/note="assembly_fragment:05139
fragment_chain:3"
misc_feature 210388..222212
/note="assembly_fragment:02948
fragment_chain:3
clone_end:77
vector_side:right"
BASE COUNT 55706 a 56624 c 54341 g 53835 t 1706 others
ORIGIN

Query Match 68.0%; Score 17; DB 2; Length 222212;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgtcaggtcaggtcag 18
|||||
Db 195872 CGTCAGGTGCGTCAG 195888

RESULT 26
AF237379 8326 bp DNA VKT 13-NOV-2000
LOCUS Xenopus laevis neurofilament protein (NF-M(2)) gene, complete cds.
DEFINITION AF237379
ACCESSION AF237379.1 GI:11139431
VERSION AF237379.1
KEYWORDS
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Fukuyoca; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodidae; Xenopus.
REFERENCE 1 (bases 1 to 8326)
Roosa,J.R., Gervasi,C. and Szaro,B.G.
Structure, biological activity of the upstream regulatory sequence,
and conserved domains of a middle molecular mass neurofilament gene
of Xenopus laevis
Brain Res. Mol. Brain Res. 82 (1-2), 35-51 (2000)
JOURNAL 11042336
PUBMED 2 (bases 1 to 8326)
REFERENCE Roosa,J.R. and Szaro,B.G.
AUTHORS Direct Submission
TITLE Submitted (18-FEB-2000) Biological Sciences, University at Albany,
JOURNAL State University of New York, 1400 Washington Avenue, Albany, NY
12222, USA

FEATURES
source
Location/Qualifiers
1..8326
/organism="Xenopus laevis"
/db_xref="taxon:8355"
TATA_signal
5'UTR 1498..1503
1524..1563
exon 1524..2598
/number=1
mRNA join(1524..2598,3533..3657,6507..>8326)
/product="neurofilament protein"
1564..>8326
gene

CDS
/gene="NF-M(2)"
join(1564..2598,3533..3657,6507..8088)
/gene="NF-M(2)"
/codon_start=1
/product="neurofilament protein"
/protein_id="AAG31697.1"
/db_xref="GI:11139432"

/translation="MSYSLETTGNPSTRRAAEITTSYSPSSSTPSSPSSPSSPSSP
TTISYKRTNLAAPRVSSNDLSFSSQSSAANGDLKQVRSSEKQSLQGLDREFGVTE
KVHLEQONKEIEAEIQALROKQSGYSQLGVNVEKREILRVNLELIGHKQAQIVLDT
EHLIEDIOLKRELEEEARIRDETAAIRAKKMEPTSTVTKVLDKQVSLQEEVAF
IKNNEEVEGELAOIOASSITVERKDEPKDIDLASALRETSOLEGSSNMLOTEPV
FNRVAKLTAEAEONKEAIRSAKDIAEYRQLSKRETSVLEGGSTGSESTPEPNDIE
ERHGHDLTSVOETIOOLNDELRTGKREMSRLREYODLVNKAALDEIAVAKRLLEG
EETRSPSSGYPTGSAIHRPTIISSTRVQARKEPRVAKVQKVEETIEETVEDD
RSDMAALANADEPAGFEAGEEEDGETOGEAEAEIYAAVESSVQAAPAE
EAGEEEREKEGEAEAEAEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE
EEBEKEEEDAEVETKAEBEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE
EGBEKEEDKDEEKEEKEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
OETAEFEETODAEOKDEGEDEFTGSGKETEKEEKEEKEEKEEKEEKEEKEEKEE
EKGEFDEGETKEAEEDAGEDKQAKSPEKSSPETISVSKAASVSSKESKQOEKES
TORDDGKEDTAVNGDVERGEKESSEKAAVYNLDSPSEDKSSGQOTVITTKI
ETTSSEGDGKTITKSVTYQKVEHEETVQEKLVSTKKEVTSHAVKEVTEDK"
2599..3532
/gene="NF-M(2)"
/number=1
3533..3657
/gene="NF-M(2)"
/number=2
3659..6507
/gene="NF-M(2)"
/number=2
6507..>8326
/gene="NF-M(2)"
/number=3
8089..>8326
/gene="NF-M(2)"
1560 c 2161 g 1974 t

BASE COUNT 2631 a 1560 c 2161 g 1974 t
ORIGIN

Query Match 64.0%; Score 16; DB 5; Length 8326;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caggtcaggtcagca 20
|||||
Db 3060 CAGGTGCGAGTCAGCA 3075

RESULT 27
AE004050 10536 bp DNA BCT 15-JUN-2001
LOCUS Xylella fastidiosa 9a5c, section 196 of 229 of the complete genome.
DEFINITION AE004050 AE003849
ACCESSION AE004050.1 GI:9107594
VERSION AE004050.1
KEYWORDS
SOURCE Xylella fastidiosa 9a5c.
ORGANISM Xylella fastidiosa 9a5c.
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.

REFERENCE 1 (bases 1 to 10536)
AUTHORS Simpson,A.J., Relnach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M., Araya,J.E., Bata,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bore,J.M., Briones,M.R.,
Bueno,M.R., Camargo,A.A., Camargo,L.F., Carraro,D.M., Carier,H.,
Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
Coutinho,L.L., Cristofani,M., Dias-Neto,B., Docena,C., El-Dorry,H.,
Pacincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
Garnier,J., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
Ho,P.L., Hohenseil,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
and Marino,C.L.

TITLE	The genome sequence of the plant pathogen <i>Xylella fastidiosa</i> . The <i>Xylella fastidiosa</i> Consortium of the Organization for Nucleotide Sequencing and Analysis
JOURNAL	Nature 406 (6792), 151-157 (2000)
MEDLINE	20365717
PUBMED	10910347
REFERENCE	2 (bases 1 to 10536)
AUTHORS	Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvaranga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Britones,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Cartaro,D.M., Carer,H., Colauto,N.B., Colombo,C., Costa,F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Pacinani,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Fromme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hobeisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,L.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani,Jr.,A., Nobrega,F.C., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira,Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa,Jr.,V.E., de Sa,Jr.,R.V., Santelli,R.V., Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terezzi,M.F., Truffi,D., Tsai,S.M., Tsunaka,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Zeldin,J. and Setubal,J.C.
TITLE	Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil
JOURNAL	
FEATURES	
source	Location/Qualifiers
gene	1. 10536
CDS	/organism="Xylella fastidiosa 9a5c"
	/db_xref="taxon:160492"
	/clone="9a5c"
	1296..2339
	/gene="XF2409"
	1296..2339
	/gene="XF2409"
	/note="similar to SP1P39369 (percent identity: 47 %/query alignment coverage: 96.8 %/subject alignment coverage: 99.4 %): identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder/Start codon shift: 2523"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAF85208.1"
	/db_xref="GI:9107595"
	/translation="MPALGYLHVGLQKSSGSRVNYEAFIAMIASHHESIOAO YGEBLRIVGLTTPETQARAIISQACRGVGEVHEKDGITVGVHALOGREYVIF SAVYSKRADGFIQDRASMLNVAASRAKNTFLVFGMDVFTAAFKSRGGLATYRK DPAVNSALCPQPLVRDLQOVSTPAVEADAEADAFLLQALNKVOREIHIYSVPWNRK IODIGAFKMOAEVEROVYVYTDQIDNTDKKIAEVLQAAARALRGVGEVHF VDRVSKVIGDDEVEFCVGSFNWFSANRSAMVAKHETSILVYRGGLADERQTRINSLR ORITDDPOOVGOA"
gene	2311..2484
CDS	/gene="XF2410"
	2311..2484
	/gene="XF2410"
	/note="hypothetical protein; identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAF85209.1"
	/db_xref="GI:9107596"
gene	/translation="MILSKSARRDAVLQRQDVDTGWQRMHAAEGHDMSPAVALLRA VALKICCLIPG"
CDS	complement(3181..5514)
	/gene="XF2411"
	complement(3181..5514)
	/gene="XF2411"
	/note="hypothetical protein; identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAF85210.1"
	/db_xref="GI:9107597"
	/translation="MLKLIKVNHHGLRNGVPTTPPGEPVTLFVANGRGKTVA NLRVTHGKDKDEVLQKARTIDPDPCSTELMDMDMDIKVTLHGMHGTDPDIL JEDATFVDNVYSGQEVYPAHAKQKLOVYLGESEKILREIYRLAKEIADRHNRKER EDRIRVYSNMVITADIRLITPTVDEAKIATLARTREVRNAPALLOGNAORPPLHC CNIDAFVSTIGTFHDIRADAEIVQAHQRQPPGIEAMVQGRNABQCGPCG QEVATATLIKAVETYEATFAKFMKNVATLGSVATRPPEACLEKRIADRIQGNACVA AKMDOLSLSPEDNPATATFTRLRTVALAALQSKORPLDCVGTQDDKADMYLLE ELRAQDAVANAIVEINRITNFRKRLNPPALDASITATLIGIVRQSEVQVAT VIOAKAKKEQLEREKKNRALLDGLDPLSEVASKINQPLRDGGAFFSTIKKOASM OGQTMKASVYDLRGEVALGRRTSDPGEFNVLSGDKRTLALFLARLAPYDAL VKKSVLDLDPKCSFDMTRRTMESIALVNGOVVYLSHDYFLRLDLRDLAPRY NKSVVWHIKRTKNDQIVSDADLDITCGSPYMLRAOVVTFVSGTEGLDEVAS ALRPLVEGFLKRFAPRLRDLSDISGMISDRKATDSDPLVLAQPYDTLEKLNAFA VQSHHDSKSLSPMNGDLROYARMALELIGCSLPH"
gene	complement(5615..5800)
CDS	/gene="XF2412"
	complement(5615..5800)
	/gene="XF2412"
	/note="hypothetical protein; identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAF85211.1"
	/db_xref="GI:9107598"
	/translation="MIFGATQTNADRRRSGTRHNAVCTGMAKRSRLRPPQVYHNHLR KPHVKVSISSPIRTGTL"
gene	6084..6512
CDS	/gene="XF2413"
	6084..6512
	/gene="XF2413"
	/note="hypothetical protein; identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAF85212.1"
	/db_xref="GI:9107599"
	/translation="HWGQVPTADPTAAAGACSEPSIKNIALSRGVHLLCAGRLM YGGCGLGQNPMPKGSVSKSVVLVGLKGSQGVQVBAARAPARTFGKYSALML YRWMCSERHCLSRTEMGATNSGDAGPRGRGMLGKRG"
gene	complement(6509..8314)
CDS	/gene="XF2414"
	complement(6509..8314)
	/gene="XF2414"
	/note="hypothetical protein; identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAF85213.1"
	/db_xref="GI:9107600"
	/translation="MPPKRHOVDAADTDKENLTVTTEHCFGVCFPIIDSGSGQIFRG FSEFLGLAMVLAWTIDVRYRFRQVAPLPLRMSFYVVAIGLTLTDLWAEQ MWVRGILITLTPAVMOALIGAFLLTITLTMALGFIIPPYQVYNAARPOOTLYRILN GSPHELAVINDELVSSEALVQATNYSKONGERIECEEDEODIPVNNRANALL LIAEKRLRAIVASPTRALFOATANKKYGIPIRVAQNTFHBAALNKDSFELHE TNGCALGIDTDHKEPTQAMFANTEMVETICTLFNHRDRANKKDPQWMLAXCSVLTMT LRSYAKQVGYSPVLSQAGVYIASASDLSITNEPSSADSPVLRRLAVLOFTIDA VRLIDGVEPHEVCIRACVYDSSRAIFCDRIANMIFVYISLSEVRLPOLEWKIOHN

SVNSLFFHGAFFKRPAGVAVKRLRLIYQIADMDFPWEVAKILLRFLCNALALAV
IRGMDYKSDRRFHHRLHRAVLAMTKRHEVWLHSHRRLAESCLGNHITYDAEHLRLA
YTHPAKGPICGVSYFALNPASPPPEVAASAVNA"

trna
/note="found by tRNAscan"
/product="tRNA-Thr-2"
/complement(8365..8440)
/gene="tRNA-Thr-2"

gene
/complement(8365..8440)
/gene="tRNA-Thr-2"

gene
/complement(8457..8648)
/gene="XP2415"

CDS
/complement(8457..8648)
/gene="XP2415"
/note="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA05214.1"
/db_xref="GI:9107601"
/translation="MPVRAVDPPPPILKSLCAVVAHLLOSIPSPDAILMRRAP
EAHKKSSALSRPSTAM"
/complement(8921..9871)
/gene="XP2416"
/complement(8921..9871)
/gene="XP2416"
/note="similar to GI15915671 (percent identity: 64 %/query
alignment coverage: 99.4 %/subject alignment coverage:
99.4 %) identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=1
/product="drug tolerance protein"

Query Match 64.0%; Score 16; DB 1; Length 10536;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 cgtcaggtgcaggtca 17
|||||

Db 1995 CGTcaggtgcaggtca 2010

RESULT 28
AE005241
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
PUBMED
JOURNAL
TITLE
MEDIAN
NATURE 409 (6819), 529-533 (2001)
11206551
2 (bases 1 to 12242)
Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7

TITLE
JOURNAL
FEATURES
source
Direct Submission
Submitted (22-Oct-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..12242
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
/complement(146..952)
/gene="tna"
/note="20755"
/complement(146..952)
/gene="tna"
/function="enzyme: Macromolecule degradation: Degradation
of RNA"
/note="Residues 1 to 268 of 268 are 99.25 pct identical to
residues 1 to 268 from *Escherichia coli* K-12 Strain
MG1655: B0611"
/codon_start=1
/transl_table=1
/product="RNase I, cleaves phosphodiester bond between any
two nucleotides"
/protein_id="AA054946.1"
/db_xref="GI:12513510"
/translation="MKAFRMAALVALSLPSSANVALOAKOYGFDRYVALSMQ
TGFCOSQDRNRERDECRIOTETTNADFLTYGLMPXLPKSVAAAGVDERMRMRG
CATPEPIPLPEARASRMCSSEPGTSLTETAKISEVAPGAGSCLEERYAAHGACF
GDDPDAVEGTWVRUNOELKESEAKPLADNIGKTVSRDRDAFAKSKWENKAYKL
TCQGNPAVITLQVSIKADATNAPLSANSLPDPHPNCGKTVIDKAG"
/complement(1066..2529)
/gene="ybd5"
/note="20756"
/complement(1066..2529)
/gene="ybd5"
/function="putative membrane; Not classified"
/note="Residues 1 to 487 of 487 are 100.00 pct identical
to residues 1 to 487 from *Escherichia coli* K-12
strain MG1655: B0612"
/codon_start=1
/transl_table=1
/product="putative a membrane protein"
/protein_id="AA054947.1"
/db_xref="GI:12513511"
/translation="MSLAKDNIMKLAPLVVMGVMLIPYDQAPPOAMHVAFAVAM
IVGMLEIPATATSAISFLAVTICVGSNYLLFDKLEADPAFNQKALMKWGLAGFEST
TWMVFGAFITGALGVEYSGRRIALFLVEMGRKRLTLCYATVITDILAPPTPNT
ARTGCTVEPIYIKNPLPKSPNDPSRRIRGGYGLMMNVISTSSMFTYGAAPVL
GLERVSKIAGIQLISWDLFCPLFVGVITLILAPWLSYVLRPEITHSEVATWAGDE
LKTMGALTREEMTLIGVLISGLWVGSEVINATVAGLVALMLALVAVPKDITR
VNSAMNTLVNIALVLVWANGI/TRSGFLDMFGAGTMSYHLEGFSPNATVILVAVFPAH
YLFASLSAHTATMPLVILVAGKKIPGPMEQCLILVLISGIMGCLPYATGAGVILY
CGGVKSKDQWRIAGIPIGVITYISMLLVGMPLIAMNN"
/complement(2580..3458)
/gene="clicg"
/note="20757"
/complement(2580..3458)
/gene="clicg"
/function="orf: Unknown function"
/note="Residues 1 to 292 of 292 are 98.97 pct identical to
residues 1 to 292 from *Escherichia coli* K-12 Strain
MG1655: B0613"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AA054948.1"
/db_xref="GI:12513512"
/translation="MSMPATLVTTTTLATSLIDEVALGVRAMLTENVISPPRGVLR
IRGAKHDMALDEDFHRSALAIQGLMPRTFEGASCEMAPEAVNLGIRTGACBQDM
FNATAGVNHGKISFSLIGLCAIGRLIDLPYTCSPGASFCRGLTRDREITN
NSQITAGORLYOOLGLTGAGAEAGYPIVNHAPHYLTLDQGLDPELTALDPTLL
LMAINDTNVASKSGRGGLRMVIREAQITLLKCGICITPADLDYLYNQFDECIERNLSP


```

gene
    GCSADLLITWFLAQI"
    complement(3433..3984)
    /gene="ybdy"
    /note="20758"
    complement(3433..3984)
    /gene="ybdy"
    /function="orf: Unknown function"
    /note="Residues 1 to 183 of 183 are 99.45 pct identical to
    residues 1 to 183 of 183 from Escherichia coli K-12 Strain
    MG1655: B0614"
    /codon_start=1
    /transl_table=11
    /product="orf, hypothetical protein"
    /protein_id="AAG54949.1"
    /db_xref="GI:12513513"
    /translation="MHLPELASHAAVSIPELLVSRDERQARQHWLKRHPVPLVFT
    VVAPGPKDSEVTRIFNKGVTALAKAQWQIQEQALVASGPEMLSTIAPAR
    DIKATIELEHNLGRWDIVLPEGEILSRDYSIPPRCLCEQSAACVARGKT
    HOLTLNRMELINDVACNVN"
    complement(3988..5520)
    /gene="citF"
    /note="20759"
    complement(3988..5520)
    /gene="citF"
    /function="enzyme: Central intermediary metabolism: Pool,
    multipurpose conversions of intermed. met'm"
    /note="Residues 1 to 510 of 510 are 99.41 pct identical to
    residues 1 to 510 of 510 from Escherichia coli K-12 Strain
    MG1655: B0615"
    /codon_start=1
    /transl_table=11
    /product="citrate lyase alpha chain"
    /protein_id="AAG54950.1"
    /db_xref="GI:12513514"
    /translation="MTQKIEQSQREQVAAVNRACEDLAAQNSFKQTYQAEKARD
    KLCANLEEAIRRSQDQMTVFHFHAFRGSDLLVNMVMDIARMKFNLLASSLSLSD
    CHAPLEHIEGQVVRITYSGLPLAEISRLAEVQIISHGGRVHLVSGELNI
    DVAFILGVSDEFNGANGYTKACGSLGAVMDANAKQVAVLTELPHNPASI
    EQQVQVLVYKPRVGDAAKIGACATPMTNPRLLIARSAADYIVNSGYEKEFSMT
    GTGGASLAVTLFLEDRMSRDLRADPAGCITATMYDLHEKILIRKLIVQSFDSIAA
    OSLARNPNHETISANQYAMGSKGASVDRIDVAVLSALTDQFNANVLGSDGVLAR
    ASGHCDTIAASLITIVAPLVGRIPITLVNLTITFGSSVDILVTDHGIANPAR
    PELAEFLQEGIKVVISIEMLRERARLLTGPQIEFTDRVAVVVRGDSVIDVHVQ
    KE"
    complement(5531..6454)
    /gene="citE"
    /note="20760"
    complement(5531..6454)
    /gene="citE"
    /function="enzyme: Central intermediary metabolism: Pool,
    multipurpose conversions of intermed. met'm"
    /note="Residues 1 to 307 of 307 are 99.34 pct identical to
    residues 1 to 307 of 307 from Escherichia coli K-12 Strain
    MG1655: B0616"
    /codon_start=1
    /transl_table=11
    /product="citrate lyase beta chain (acyl lyase subunit)"
    /protein_id="AAG54951.1"
    /db_xref="GI:12513515"
    /translation="MGLPMISASLQQRKTRPRSMFVPGANAANVSFTYPADAL
    MPDEESVALREKDTARRMYHALQPLVDITIVRVNADSEMGVNDLEAVRGCA
    DVARLPKTDADVDLIEKILIRKAKGSPSTGLIAIESPLGITAVETVAHSE
    RLGIALGADYVRNLRTSPSPGTELLPARCSLIDASAGIQADITYSDNNAG
    FIDPAHIKQIGFDGKSLNPROIDLHNIYAFQKEVDHARVDAEAARREGICV
    VSLNGKWDVPIIDRARLVXSRALSGIREE"
    complement(6436..6732)
    /gene="citD"
    /note="20761"
    complement(6436..6732)
    /gene="citD"
    /function="enzyme: Central intermediary metabolism: Pool,
    multipurpose conversions of intermed. met'm"
    /note="Residues 1 to 98 of 98 are 100.00 pct identical to

```

```

residues 1 to 98 of 98 from Escherichia coli K-12 Strain
MG1655: B0617"
/codon_start=1
/transl_table=11
/product="citrate lyase acyl carrier protein (gamma
chain)"
/protein_id="AAG54952.1"
/db_xref="GI:12513516"
/translation="MKINQPAVAGTLESGDWMIRIAPLDQIDIDQINSVKEQFGDA
IRTTILDVLYARNVKGVLNVDDKGDALDCILRARLEALRLARASITRALPREDQ"
complement(6747..7892)
/gene="citC"
/note="20762"
complement(6747..7892)
/gene="citC"
/translation="orf: Unknown function"
/translation="enzyme: Central intermediary metabolism: Pool,
multipurpose conversions of intermed. met'm"
/note="Residues 1 to 381 of 381 are 99.21 pct identical to
residues 1 to 381 of 381 from Escherichia coli K-12 Strain
MG1655: B0618"
/codon_start=1
/transl_table=11
/product="citrate lyase synthetase (citrate (pro-3S)-lyase
ligase"
/protein_id="AAG54953.1"
/db_xref="GI:12513517"

```

Query Match 64.0%; Score 16; DB 1; Length 12242;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 9 tgaggtcagccactt 24
Db 4236 TGCAGGTCAACACGTT 4251
|||||

```

RESULT 29
 LOCUS MLCB1450 38065 bp DNA BCF 27-AUG-1999
 DEFINITION Mycobacterium leprae cosmid B1450.
 ACCESSION AL035159
 VERSION 60
 KEYWORDS 60 kD chaperonin; CDP-diacylglycerol--serine
 O-phosphatidyltransferase; cpn60; glutamine-binding protein;
 groEL-2; hydrolase; oxidoreductase; phosphomethylpyrimidine kinase;
 pseudogene; psbA; RLEP1; serine/threonine protein kinase; thiamine
 biosynthesis protein ThIG; thiamine-phosphate pyrophosphorylase;
 thID; thIE; thIG.
 SOURCE Mycobacterium leprae.
 ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
 Mycobacterium.
 1 (bases 1 to 38065)
 REFERENCE 1
 AUTHORS Eigmeier, K., Honore, N., Woods, S. A., Caudron, B. and Cole, S. T.
 TITLE Use of an ordered cosmid library to deduce the genomic organization
 of Mycobacterium leprae
 JOURNAL Mol. Microbiol. 7 (2), 197-206 (1993)
 MEDLINE 93188700
 REFERENCE 2 (bases 1 to 38065)
 AUTHORS Brown, D. and Churcher, C. M.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 38065)
 AUTHORS James, K. D., Parkhill, J., Barrell, B. G. and Rajandream, M. A.
 TITLE Direct Submission
 SUBMITTED (07-JAN-1998) Mycobacterium leprae sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridg
 CB10 1SA E-mail: barreli@sanger.ac.uk Cosmids supplied by Dr.
 Stewart T. Cole. [3] Unite de Genetique Moleculaire Bacterienne,
 Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
 France Requests for cosmids should be sent to Karin Eigmeier
 (keigmeier@fr)

COMMENT

/gene="MLCB1450.06c"
 /note="MLCB1450.06c, probable transcriptional regulator,
 len: 189 aa; similar to hypothetical proteins from M.
 tuberculosis and S. coelicolor, and equivalent to
 M. tuberculosis putative transcriptional regulator Rv0067c
 (68.8% identity in 189 aa overlap). Contains probable
 helix-turn-helix motif at aa 35-56 (Score 1794, +5.30 SD)
 and Pfam match to entry PF00440 tetr, Bacterial regulatory
 proteins, tetr family"
 /codon_start=1
 /transl_table=11
 /product="putative transcriptional regulator"
 /protein_id="CA22690.1"
 /db_xref="GI:4154041"
 /db_xref="SPTREMBL:O92BM6"
 /translation="MAQAPAPLRADARNRARILGVAYENAFATGLSPIDETARRAG
 VGGIVYRHPPTKALCAAVIGDRPHLVDDGYLLKSAGPGALFTYLRSLVHMGCA
 TDRGLVDALAGAGGIDILGVAADPAFLITLSDLLYAQYAGTARTDVGVRKSLV
 VGCAMKAVNSALAEKRTDYYVDCGLRAR"

Query Match 64.0%; Score 16; DB 1; Length 38065;

Best Local Similarity 100.0%; Pred. No. 80;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caggtgcaggtcagca 20

|||||

Db 14193 CAGGTGCAGGTGACGA 14208

RESULT 30

AC091120

LOCUS

DEFINITION

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

AC091120

Query Match 64.0%; Score 16; DB 2; Length 39520;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caggtgcaggtcagca 20

|||||

Db 1027 CAGGTGCAGGTGACGA 1042

RESULT 31

AC004508

LOCUS

DEFINITION

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

AC004508

```

repeat_region complement(6153, .6473)
/rpt_family="Alu"
repeat_region complement(6485, .6565)
/rpt_family="MER44C"
repeat_region complement(6779, .6989)
/rpt_family="MER46"
repeat_region complement(7235, .8312)
/rpt_family="L1"
repeat_region complement(8410, .8569)
/rpt_family="L1"
repeat_region 8577, .8987
/rpt_family="THE1"
repeat_region complement(9020, .9511)
/rpt_family="L1"
repeat_region 9872, .10296
/rpt_family="THE1"
repeat_region 10297, .11909
/rpt_family="MSTAR"
repeat_region 11929, .12309
/rpt_family="THE1"
repeat_region 12868, .13131
/rpt_family="Alu"
repeat_region complement(13267, .13635)
/rpt_family="L1"
prim_transcript join(14992, .15122, 21091, .21151, 22200, .22263, 25113, .25251, 33644, .33721)
/misc_feature /standard_name="y99d04"
/note="98% identity R74188"
14993, .15116
/note="GRAIL 2 excellent exon, frame 0"
prim_transcript join(15038, .15122, 21091, .21151, 22209, .22263, 25113, .25251, 33644, .33885)
/standard_name="z055b03"
/note="98% identity AA402132"
repeat_region complement(15991, .16164)
/rpt_family="MER42"
repeat_region complement(16400, .16743)
/rpt_family="MLT2B2"
repeat_region complement(16761, .17073)
/rpt_family="Alu"
repeat_region 16768, .16787
/note="(T) 20"
/rpt_type=tandem
/rpt_unit=T
misc_feature 19581, .19626
/note="GRAIL 2 excellent exon, frame 0"
20435, .20765
/rpt_family="Alu"
repeat_region 23936, .24290
/rpt_family="Alu"
24273, .24292
/note="(A) 20"
/rpt_type=tandem
/rpt_unit=A
misc_feature 25115, .25247
/note="GRAIL 2 excellent exon, frame 2"
26872, .26965
/rpt_family="MTR"
repeat_region 29019, .29360
/rpt_family="Alu"
repeat_region 29356, .31388
/rpt_family="L1"
repeat_region 31440, .32072
/rpt_family="L1"
repeat_region complement(32120, .33230)
/rpt_family="L1"
32581, .32624
/note="(AC) 22"
/rpt_type=tandem
/rpt_unit=AC
repeat_region complement(32668, .32776)
/rpt_family="MER3"
misc_feature 32932, .33120

```

```

repeat_region /note="96% identity zh52f10.r1"
/db_xref="dbEST:M78940"
complement(33133, .33199)
/rpt_family="MER3"
repeat_region complement(33644, .33883)
/note="99% identity n66d01.s1"
/db_xref="dbEST:AA515600"
33648, .33737
/note="GRAIL 2 excellent exon, frame 2"
35260, .35660
/rpt_family="MLT1"
repeat_region 36246, .36533
/rpt_family="Alu"
repeat_region complement(39299, .39589)
/rpt_family="Alu"
repeat_region complement(39781, .39931)
/rpt_family="Alu"
repeat_region complement(41134, .41588)
/rpt_family="MLT2"
41621, .41706
/rpt_family="MLT1"
41748, .41771
/note="(GTT) 6"
/rpt_type=tandem
/rpt_unit=GTT
repeat_region complement(41757, .42397)
/rpt_family="L1"
42935, .43106
/rpt_family="Alu"
complement(44760, .44802)
/note="GRAIL 2 excellent exon, frame 2"
complement(49449, .49589)
/rpt_family="MIR2"
repeat_region complement(51869, .52337)
/rpt_family="MLT1"
52773, .53082
/rpt_family="Alu"
<53503, .>67215
/note="MX2"
/note="interferon-regulated resistance GTP-binding protein
MXB (P78-related protein)"
/evidence=experimental
join(<53503, .53668, 54302, .54494, 56530, .56664, 58923, .59077, 67077, .>67215)
/gene="MX2"
/note="interferon-regulated resistance GTP-binding protein
MXB (P78-related protein)"
/codon_start=2
/evidence=experimental
/product="MX2"
/protein_id="AAC08448.1"
/db_xref="GI:2996631"
/translacion="FOQPPPEFGVPPQMMPPNNQGAEDAAFLAKDPNFTLLNNO
PGRNSQPRAMGPNENLYSOYBOKVPICIDLSIALGVEDDLPAIVAGQSSG
KSSVLEALSGVALPRGSGIVTRGCLVTKKQPCAMGRISYRTLELDDPGVEK
ETIKKQNMWAGNKGISHELISLETSPEVPDUTITDLPGLTRAVDQNPDRIGLQIK
ALKKTIQROQITNLVWPCNVNDIATTEALSMARHVDPEGRIT"
57922, .58215
/rpt_family="Alu"
59599, .59877
/rpt_family="Alu"
60428, .60810
/rpt_family="THE1"
61246, .61366
/rpt_family="Alu"

```

```

Query Match 64.0% Score 16; DB 9; Length 71230;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 gtaagtgcaagtcag 18
Db 40455 GTCAGGTGTCAGTCAG 40470

```


	repeat_region	/rpt_family="L1"	40345. .40893	
		/rpt_family="L1"	complement(41771. .47939)	
	repeat_region	/rpt_family="L1"	complement(47950. .48026)	
		/rpt_family="MER"	complement(48181. .48428)	
	repeat_region	/rpt_family="MER"	48233. .48439	
		/rpt_family="MER"	complement(48619. .48655)	
	repeat_region	/rpt_family="MER"	complement(48694. .48890)	
		/rpt_family="MER"	48819. .48974	
	repeat_region	/rpt_family="MER"	49030. .49292	
		/rpt_family="MER"	49440. .49757	
	repeat_region	/rpt_family="MER"	complement(49570. .49873)	
		/rpt_family="MER"	51098. .51155	
	repeat_region	/rpt_family="ALU"	51181. .51230	
		/rpt_family="ALU"	complement(51563. .51830)	
	repeat_region	/rpt_family="ALU"	complement(52400. .52441)	
		/rpt_family="L1"	complement(53064. .53085)	
	repeat_region	/rpt_family="L1"	complement(53087. .53384)	
		/rpt_family="ALU"	complement(53410. .53435)	
	repeat_region	/rpt_family="L1"	complement(53436. .53545)	
		/rpt_family="ALU"	complement(53554. .53709)	
	repeat_region	/rpt_family="ALU"	complement(60972. .60992)	
		/rpt_family="L1"	complement(64802. .64824)	
	repeat_region	/rpt_family="L1"	complement(66753. .67258)	
		/rpt_family="MER"	complement(72167. .72185)	
	repeat_region	/rpt_family="L1"	72879. .73153	
		/rpt_family="ALU"	complement(82252. .82468)	
	repeat_region	/rpt_family="MER"	82281. .82473	
		/rpt_family="MER"	82600. .82926	
	repeat_region	/rpt_family="MER"	complement(83480. .83780)	
Query Match				
Best Local Similarity	64.0%; Score 16; DB 9;		length 85633;	
Matches	16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
OY	5 caggtgcaggtcagca 20			
DB	22093 CAGGTGCAGGTCACGA 22078			
RESULT 33				
AC007555/c	AC007555 109973 bp DNA		HTG 24-AUG-2000	
LOCUS	Homo sapiens chromosome 7 clone RP5-925B6, WORKING DRAFT SEQUENCE,			
DEFINITION	2 unordered pieces.			

```

AC007555
AC007555.2 GI:9887766
HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 109973)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 109973)
Waterston,R.H.
Direct Submission
Submitted (14-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis
MO 63108, USA
On Aug 24, 2000 this sequence version replaced gi:4827326.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.D0925B06
----- Summary Statistics -----
Sequencing vector: M13; 79%
Sequencing vector: plasmid; 21%
Chemistry: Dye-Primer ET; 79% of reads
Chemistry: Dye-terminator Big Dye; 21% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 109177 bases at least Q40
Consensus quality: 109655 bases at least Q40
Consensus quality: 109773 bases at least Q20
Insert size: 201000; agarose-1p
Insert size: 0; sum-of-contigs
Quality coverage: in Q20 bases; agarose-1p
Quality coverage: in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 25831: contig of 25831 bp in length
* 25832 25931: gap of unknown length
* 25932 109973: contig of 84042 bp in length.
FEATURES
SOURCE
1. 109973
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP5-925B6"
1. 25831
/note="assembly_name:Contig5
clone_end:SP6
vector_side:left"
25932..109973
/note="assembly_name:Contig6
clone_end:77
vector_side:left"
BASE COUNT 34650 a 19714 c 19631 g 35877 t 101 others
ORIGIN
Query Match 64.0%; Score 16; DB 2; Length 109973;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

|||||
db 91300 CAGTGCAGTCAGCA 91285

RESULT 34
AC092845/c
LOCUS
DEFINITION
AC092845 119749 bp DNA HTG 30-JUL-2001
Homo sapiens chromosome UNK clone RP11-480J15, *** SEQUENCING IN
PROGRESS ***, 46 unordered pieces.
AC092845
VERSION
AC092845.1 GI:15029484
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 119749)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 119749)
Waterston,R.H.
Direct Submission
Submitted (30-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_MH0480J15

* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1171: contig of 1171 bp in length
1172 1271: gap of unknown length
1272 2890: contig of 1619 bp in length
2891 2990: gap of unknown length
2991 4472: contig of 1482 bp in length
4473 4572: gap of unknown length
4573 5850: contig of 1278 bp in length
5851 5950: gap of unknown length
5951 7292: contig of 1342 bp in length
7293 7392: gap of unknown length
7393 9328: contig of 1936 bp in length
9329 9428: gap of unknown length
9429 11085: contig of 1657 bp in length
11086 11185: gap of unknown length
11186 12927: contig of 1742 bp in length
12928 13027: gap of unknown length
13028 14570: contig of 1543 bp in length
14571 14670: gap of unknown length
14671 15742: contig of 1072 bp in length
15743 15842: gap of unknown length
15843 17083: contig of 1241 bp in length
17084 17183: gap of unknown length
17184 18697: contig of 1514 bp in length
18698 18797: gap of unknown length
18798 19875: contig of 1078 bp in length
19876 21390: contig of 1415 bp in length
21391 21490: gap of unknown length
21491 23170: contig of 1680 bp in length
23171 23270: gap of unknown length
23271 24798: contig of 1528 bp in length

```

```

* 24799 24898: gap of unknown length
* 24899 27682: contig of 2784 bp in length
* 27683 27782: gap of unknown length
* 27783 29997: contig of 2215 bp in length
* 29998 30097: gap of unknown length
* 30098 32119: contig of 2022 bp in length
* 32120 32219: gap of unknown length
* 32220 34076: contig of 1856 bp in length
* 34076 34175: gap of unknown length
* 34176 36111: gap of unknown length
* 36112 36711: gap of unknown length
* 36712 38912: contig of 2201 bp in length
* 38913 39012: gap of unknown length
* 39013 40658: contig of 1646 bp in length
* 40659 40759: gap of unknown length
* 40759 42886: contig of 2127 bp in length
* 42886 42985: gap of unknown length
* 42986 44675: gap of 1690 bp in length
* 44676 44775: contig of 1690 bp in length
* 44776 46101: gap of unknown length
* 46102 46201: contig of 1326 bp in length
* 46201 46201: gap of unknown length
* 46202 47837: gap of 1636 bp in length
* 47838 47937: gap of unknown length
* 47938 50354: contig of 2417 bp in length
* 50355 50454: gap of unknown length
* 50455 53484: contig of 3030 bp in length
* 53485 53584: gap of unknown length
* 53585 57512: contig of 3928 bp in length
* 57513 57612: gap of unknown length
* 57613 59689: contig of 2077 bp in length
* 59690 59789: gap of unknown length
* 59790 62280: contig of 2491 bp in length
* 62281 62380: gap of unknown length
* 62381 65665: contig of 3285 bp in length
* 65666 65765: gap of unknown length
* 65766 68813: contig of 3048 bp in length
* 68814 68913: gap of unknown length
* 68914 72349: contig of 3436 bp in length
* 72350 72449: gap of unknown length
* 72450 74667: contig of 2418 bp in length
* 74668 74967: gap of unknown length
* 74968 77456: gap of 2489 bp in length
* 77457 77556: gap of unknown length
* 77557 80059: contig of 2503 bp in length
* 80060 80159: gap of unknown length
* 80160 82473: contig of 2314 bp in length
* 82474 82574: gap of unknown length
* 82574 86092: contig of 3519 bp in length
* 86093 86192: gap of unknown length
* 86193 89565: contig of 3373 bp in length
* 89566 89665: gap of unknown length
* 89666 92631: contig of 2966 bp in length
* 92632 92732: gap of unknown length
* 92733 96074: contig of 3343 bp in length
* 96075 96174: gap of unknown length
* 96175 100717: contig of 4543 bp in length
* 100718 100817: gap of unknown length
* 100818 107308: contig of 6491 bp in length
* 107309 107409: gap of unknown length
* 107409 119749: contig of 12341 bp in length.

FEATURES
source
1..119749
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-480J15"
1..1171
/note="assembly_name:Contig11"
1272..2890
/note="assembly_name:Contig27"
2991..4472
/note="assembly_name:Contig32"
4573..5850
misc_feature
misc_feature
misc_feature
misc_feature

```

```

misc_feature      /note="assembly_name:Contig35"
                    5951..7292
                    /note="assembly_name:Contig36"
misc_feature      /note="assembly_name:Contig37"
                    7393..9328
                    /note="assembly_name:Contig38"
misc_feature      /note="assembly_name:Contig39"
                    9429..11085
                    /note="assembly_name:Contig40"
misc_feature      /note="assembly_name:Contig41"
                    1186..12927
                    /note="assembly_name:Contig42"
misc_feature      /note="assembly_name:Contig43"
                    13028..14570
                    /note="assembly_name:Contig44"
misc_feature      /note="assembly_name:Contig45"
                    14671..15742
                    /note="assembly_name:Contig46"
misc_feature      /note="assembly_name:Contig47"
                    15843..17083
                    /note="assembly_name:Contig48"
misc_feature      /note="assembly_name:Contig49"
                    17184..18697
                    /note="assembly_name:Contig50"
misc_feature      /note="assembly_name:Contig51"
                    18798..19875
                    /note="assembly_name:Contig52"
misc_feature      /note="assembly_name:Contig53"
                    19976..21390
                    /note="assembly_name:Contig54"
misc_feature      /note="assembly_name:Contig55"
                    21491..23170
                    /note="assembly_name:Contig56"
misc_feature      /note="assembly_name:Contig57"
                    23271..24798
                    /note="assembly_name:Contig58"
misc_feature      /note="assembly_name:Contig59"
                    24899..27682
                    /note="assembly_name:Contig60"
misc_feature      /note="assembly_name:Contig61"
                    27783..29997
                    /note="assembly_name:Contig62"
misc_feature      /note="assembly_name:Contig63"
                    30098..32119
                    /note="assembly_name:Contig64"
misc_feature      /note="assembly_name:Contig65"
                    32220..34075
                    /note="assembly_name:Contig66"
misc_feature      /note="assembly_name:Contig67"
                    34176..36611
                    /note="assembly_name:Contig68"
misc_feature      /note="assembly_name:Contig69"
                    36712..38912
                    /note="assembly_name:Contig70"
misc_feature      /note="assembly_name:Contig71"
                    39013..40658
                    /note="assembly_name:Contig72"
misc_feature      /note="assembly_name:Contig73"
                    40759..42885
                    /note="assembly_name:Contig74"
misc_feature      /note="assembly_name:Contig75"
                    42986..44675
                    /note="assembly_name:Contig76"
misc_feature      /note="assembly_name:Contig77"
                    44776..46101
                    /note="assembly_name:Contig78"
misc_feature      /note="assembly_name:Contig79"
                    46202..47837
                    /note="assembly_name:Contig80"
misc_feature      /note="assembly_name:Contig81"
                    47938..50354
                    /note="assembly_name:Contig82"
misc_feature      /note="assembly_name:Contig83"
                    50455..53484
                    /note="assembly_name:Contig84"
misc_feature      /note="assembly_name:Contig85"
                    53585..57512
                    /note="assembly_name:Contig86"
misc_feature      /note="assembly_name:Contig87"
                    57613..59689
                    /note="assembly_name:Contig88"
misc_feature      /note="assembly_name:Contig89"
                    59790..62280
                    /note="assembly_name:Contig90"
misc_feature      /note="assembly_name:Contig91"
                    62381..65665

```

Query Match 64.0%; Score 16; DB 2; Length 119749;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caggtcaggtcagca 20
 |||
 Db 930 CAGGTGACGTCAGCA 915

RESULT 35
 AC073052/c
 LOCUS AC073052 127070 bp DNA PRI 09-AUG-2001

```

DEFINITION Homo sapiens chromosome 2 clone RP11-92F20, complete sequence.
ACCESSION AC073052
VERSION AC073052.5 GI:15145540
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
          1 (bases 1 to 127070)
          Waterston, R.H.
          The sequence of Homo sapiens clone
          2 (bases 1 to 127070)
          Waterston, R.H.
          Submitted (08-JUN-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          3 (bases 1 to 127070)
          Waterston, R.H.
          Direct Submission
          Submitted (09-AUG-2001) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          On Aug 9, 2001 this sequence version replaced gi:14249098.
COMMENT
REFERENCE AUTHORS
          DIRECT SUBMISSION
          JOURNAL

```

FEATURES

source

```

1..127070
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="2"
  /clone="RP11-92F20"
BASE COUNT 41948 a 25328 c 21891 g 37903 t
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 9; Length 127070;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 caggtcaggtcagca 20
  |||
  Db 124378 CAGGTGACGTCAGCA 124363

```

```

RESULT 36
AC027810/c
LOCUS AC027810 155628 bp DNA HMG 26-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-357J12 map 18, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
ACCESSION AC027810
VERSION AC027810.3 GI:8082020
KEYWORDS HMG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
          1 (bases 1 to 155628)
          Birren, B., Linton, L., Nusbaum, C. and Lander, E.
          Homo sapiens chromosome 18, clone RP11-357J12
          2 (bases 1 to 155628)
          Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
          Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedd, F.,
          Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,

```


TITLE
JOURNAL
COMMENT

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferrelle, P., Fitzhugh, W., Gage, D.,
Gajagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Harford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Labroque, K., Lamazares, R., Landers, T., Lehotzky, J.,
Levine, R., Liu, G., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mienna, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rottman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced y1:7684514.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 19097

Center clone name: 357_J_12

Sequencing Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 146315 bases at least Q40

Consensus quality: 153082 bases at least Q20

Insert size: 163000; agarose-1p

Insert size: 154328; sum-of-ctrls

Quality coverage: 4.2 in Q20 bases; agarose-1p

Quality coverage: 4.4 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently

* consists of 14 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1068: contig of 1068 bp in length

1069 1168: gap of 100 bp

1169 2717: contig of 1549 bp in length

2718 2817: gap of 100 bp

2818 6040: contig of 3223 bp in length

6041 6140: gap of 100 bp

6141 10201: contig of 4061 bp in length

10202 10301: gap of 100 bp

10302 14615: contig of 4314 bp in length

14616 14715: gap of 100 bp

14716 21086: contig of 6371 bp in length

21087 21186: gap of 100 bp

21187 25767: contig of 4581 bp in length

25768 25867: gap of 100 bp

25868 32500: contig of 6633 bp in length

32501 32600: gap of 100 bp

32601 38651: contig of 6051 bp in length

38652 38751: gap of 100 bp

38752 50922: contig of 12171 bp in length

50923 51022: gap of 100 bp

FEATURES
Source
* 51023 71014: contig of 19992 bp in length
* 71015 71114: gap of 100 bp
* 71115 93563: contig of 22449 bp in length
* 93564 93663: gap of 100 bp
* 93664 123080: contig of 29417 bp in length
* 123081 123180: gap of 100 bp
* 123181 155628: contig of 32448 bp in length.
Location/Qualifiers
1. 155628 "Homo sapiens"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-357J12"
/clone_11b="RPCT-11 Human Male BAC"
1. 1068
/note="assembly-fragment"
1169. 2717
/note="assembly-fragment"
2818. 6040
/note="assembly-fragment"
6141. 10201
/note="assembly-fragment"
10302. 14615
/note="assembly-fragment"
14716. 21086
/note="assembly-fragment"
21187. 25767
/note="assembly-fragment"
25868. 32500
/note="assembly-fragment"
32601. 38651
/note="assembly-fragment"
38752. 50922
/note="assembly-fragment"
51023. 71014
/note="assembly-fragment"
71115. 93563
/note="assembly-fragment"
93664. 123080
/note="assembly-fragment"
123181. 155628
/note="assembly-fragment"
clone_end:SP6
vector_side:left

BASE COUNT 37655 a 38986 c 40608 g 37067 t 1302 others
ORIGIN
Query Match 64.0%; Score 16; DB 2; Length 155628;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gtcaggtcagtcag 18
DB 95800 gtcaggtcagtcag 95785

RESULT 37
AC073247
LOCUS Homo sapiens chromosome 15 clone CTD-2006N8, WORKING DRAFT
DEFINITION AC073247
ACCESSION AC073247.2 GI:8571764
VERSION AC073247.2 GI:8571764
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 156119)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Unpublished
2 (bases 1 to 156119)
Waterston, R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 18, 2000 this sequence version replaced g1:8468936.

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Project Information
Center project name: H_MS2006N08
Summary Statistics

Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 139572 bases at least Q40
Consensus quality: 144812 bases at least Q30
Consensus quality: 147446 bases at least Q20
Insert size: 160000; agarose-fp
Insert size: 154019; sum-of-coverage
Quality coverage: 4.17 in Q20 bases; agarose-fp
Quality coverage: 4.44 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently
consists of 22 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1286: contig of 1286 bp in length
1287 1386: gap of unknown length
1387 3603: contig of 2217 bp in length
3604 3703: gap of unknown length
3704 6043: contig of 2340 bp in length
6044 6143: gap of unknown length
6144 9355: contig of 3211 bp in length
9355 9454: gap of unknown length
9455 12077: contig of 2623 bp in length
12078 12177: gap of unknown length
12178 15985: contig of 3808 bp in length
15986 16085: gap of unknown length
16086 19801: contig of 3716 bp in length
19802 25990: contig of 6089 bp in length
25991 26090: gap of unknown length
26091 30598: contig of 4508 bp in length
30599 30698: gap of unknown length
30699 36263: contig of 5565 bp in length
36264 36363: gap of unknown length
36364 42783: contig of 6420 bp in length
42784 42883: gap of unknown length
42884 49479: contig of 6595 bp in length
49479 49578: gap of unknown length
49578 56751: contig of 7173 bp in length
56752 56851: gap of unknown length
56852 61885: contig of 5034 bp in length
61886 61985: gap of unknown length
61986 69384: contig of 7399 bp in length
69385 69484: gap of unknown length
69485 76911: contig of 7427 bp in length
76912 77012: gap of unknown length
77012 86274: contig of 9263 bp in length
86275 86375: gap of unknown length
86375 96575: contig of 10201 bp in length
96576 96675: gap of unknown length

FEATURES
source
1. 156119
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/clone="CTD-2006N8"
BASE COUNT 42637 a 33200 c 33979 g 44179 t 2124 others

Query Match 64.0%; Score 16; DB 2; Length 156119;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caagtcagtcagca 20
|||||
Db 90601 CAGTCAGTCAGCA 90616

RESULT 38

AC040930

LOCUS

AC040930 162043 bp DNA HTG 11-JUN-2001
DEFINITION Homo sapiens chromosome 15 clone CTD-2006N8 map 15, WORKING DRAFT

SEQUENCE, 11 unordered pieces.

AC040930.3 GI:12658160

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 162043)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 15, clone CTD-2006N8

REFERENCE

JOURNAL

AUTHORS

Unpublished

2 (bases 1 to 162043)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,

Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

Collamore, A., Cooke, P., DeArillano, R., Dewar, K., Diaz, J.S.,

Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,

Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Hooton, L.,

Howland, D.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., Lakoque, K., Lamazares, R., Landers, J., Lehoczy, J.,

Levine, R., Liu, G., Liu, G., Locke, K., MacDonald, P., Marguis, N.,

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,

Meldrum, J., Meneses, L., Mihova, T., Miranda, C., Mienda, V., Morrow, J.,

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Piere, N.,

Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tessier, S., Theodore, J., Tirrell, A., Travers, M., Triggilio, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 3, 2001 this sequence version replaced g1:7705184.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/kv/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

```

misc_feature      48697..76612
                  /note="assembly_fragment"
misc_feature      76713..162043
                  /note="assembly_fragment"
BASE COUNT      44171 a 35675 c 34546 g 46650 t      1001 others
ORIGIN
Query Match      64.0%; Score 16; DB 2; Length 162043;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      5 caagtcagtcacgaqa 20
          |||||c|||c|||c|||
Db      82976 CAGTCGACGTCACGA 82991

RESULT 39
LOCUS      AC068538/c      13-AUG-2000
DEFINITION      Homo sapiens chromosome 2 clone RP11-14164, WORKING DRAFT SEQUENCE.
ACCESSION      AC068538
VERSION      AC068538.5 GI:9795677
KEYWORDS      HTG_HTS.PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 175459)
TITLE      Waterston,R.H.
AUTHORS      The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 175459)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (03-MAY-2000) Genome Sequencing Center, Washington
REFERENCE      MO 63108 USA
AUTHORS      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
TITLE      On Aug 11, 2000 this sequence version replaced gi:9690368.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0141G04
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158616 bases at least Q40
Consensus quality: 162423 bases at least Q30
Consensus quality: 164318 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 172659; sum-of-contigs
Quality coverage: 4.51 in Q20 bases; agarose-fp
Quality coverage: 4.87 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1      1355: contig of 1355 bp in length
*      1356      1455: gap of unknown length
*      1456      3082: contig of 1627 bp in length
*      3083      3182: gap of unknown length

```

```

* 3183 4263: contig of 1081 bp in length
* 4264 4363: gap of unknown length
* 4364 5626: contig of 1263 bp in length
* 5627 5727: gap of unknown length
* 5728 7534: contig of 1807 bp in length
* 7535 7634: gap of unknown length
* 7635 9203: contig of 1569 bp in length
* 9204 9303: gap of unknown length
* 9304 11061: contig of 1759 bp in length
* 11062 14017: contig of 2856 bp in length
* 14018 14117: gap of unknown length
* 14118 16928: contig of 2811 bp in length
* 16929 17029: gap of unknown length
* 17030 19922: contig of 2894 bp in length
* 19923 20023: gap of unknown length
* 20024 23986: contig of 3964 bp in length
* 23987 24087: gap of unknown length
* 24088 27864: contig of 3778 bp in length
* 27865 27964: gap of unknown length
* 27965 31695: contig of 3731 bp in length
* 31696 31795: gap of unknown length
* 31796 34671: contig of 2876 bp in length
* 34672 34772: gap of unknown length
* 34773 39459: contig of 4688 bp in length
* 39460 39560: gap of unknown length
* 39561 43565: contig of 4006 bp in length
* 43566 43666: gap of unknown length
* 43667 47319: contig of 3653 bp in length
* 47320 53986: contig of 6568 bp in length
* 53987 54087: gap of unknown length
* 54088 61187: contig of 7100 bp in length
* 61188 61287: gap of unknown length
* 61288 66331: contig of 5045 bp in length
* 66332 66432: gap of unknown length
* 66433 72447: contig of 6016 bp in length
* 72448 72548: gap of unknown length
* 72549 80701: contig of 8154 bp in length
* 80702 80802: gap of unknown length
* 80803 89408: contig of 8607 bp in length
* 89409 89509: gap of unknown length
* 89510 100141: contig of 10632 bp in length
* 100142 100241: gap of unknown length
* 100242 109860: contig of 9620 bp in length
* 109861 109960: gap of unknown length
* 109961 120505: contig of 10545 bp in length
* 120506 135065: gap of unknown length
* 135066 135165: contig of 14460 bp in length
* 135166 153777: gap of unknown length
* 153778 153878: contig of 18612 bp in length
* 153879 175459: gap of unknown length
* 175460 175459: contig of 21582 bp in length.

FEATURES
source
    location=Qualifiers
    1..175459
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="2"
    /clone="RP11-141G4"
    1..1355
    /note="assembly_name:Contig9"
    1456..3082
    /note="assembly_name:Contig11"
    3183..4263
    /note="assembly_name:Contig13"
    4364..5626
    /note="assembly_name:Contig14"
    5727..7533
    /note="assembly_name:Contig15"
    7634..9202
    /note="assembly_name:Contig16"
    clone_end:77
    vector_side:left"

misc_feature
    9303..11061
    /note="assembly_name:Contig17"
    11162..14017
    /note="assembly_name:Contig18"
    14118..16928
    /note="assembly_name:Contig19"
    17029..19922
    /note="assembly_name:Contig20"
    20023..23986
    /note="assembly_name:Contig21"
    24087..27864
    /note="assembly_name:Contig22"
    27965..31695
    /note="assembly_name:Contig23"
    31796..34671
    /note="assembly_name:Contig24"
    34772..39459
    /note="assembly_name:Contig25"
    39560..43565
    /note="assembly_name:Contig26"
    43666..47318
    /note="assembly_name:Contig27"
    47419..53986
    /note="assembly_name:Contig28"
    54087..61186
    /note="assembly_name:Contig29"
    61287..66331
    /note="assembly_name:Contig30"
    66432..72447
    /note="assembly_name:Contig31"
    clone_end:SP6
    vector_side:right"
    72548..80701
    /note="assembly_name:Contig32"
    80802..89408
    /note="assembly_name:Contig33"
    89509..100140
    /note="assembly_name:Contig34"
    100241..109860
    /note="assembly_name:Contig35"
    109961..120505
    /note="assembly_name:Contig36"
    120606..135065
    /note="assembly_name:Contig37"
    135166..153777
    /note="assembly_name:Contig38"
    153878..175459
    /note="assembly_name:Contig39"

BASE COUNT 47828 a 38505 c 38363 g 47961 t 2802 others
ORIGIN

Query Match 64.0%; Score 16; DB 2; Length 175459;
Best local similarity 100.0%; Pred. No. 62;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagca 20
Db 79767 CAGGTGCAGGTCA 79752

RESULT 40
AL139326/c
LOCUS
DEFINITION Human DNA sequence from clone RP11-351K3 on chromosome 13, complete
ACCESSION AL139326
VERSION AL139326.15 GI:14018249
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```

REFERENCE 1 (bases 1 to 175588)
 AUTHORS Blakey, S.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humphrey@sanger.ac.uk
 Requests: clonerequests@sanger.ac.uk

COMMENT

On May 11, 2001 this sequence version replaced gi:13277016.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats: all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TrEMBL, Wp: WormPEP; information on the WormPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
 Rpl1-351k3 is from the library RPl1-11.2 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBCEc3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RPl1-351k3. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RPl1-351k3 is at 1 in this sequence. The
 true left end of clone RPl1-154C3 is at 175489 in this sequence.

FEATURES

source

1..175588

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="13"

/clone="Rpl1-351k3"

/clone_lib="RPl1-11.2"

68..360

/note="AluX repeat: matches 1..290 of consensus"

953..1089

/note="MIR repeat: matches 80..210 of consensus"

3306..3437

/note="L1M1 repeat: matches 6066..6201 of consensus"

3591..3719

/note="MER47 repeat: matches 1..133 of consensus"

3722..3797

/note="MER47 repeat: matches 2248..2323 of consensus"

3803..3878

/note="L1M2 repeat: matches 2203..2279 of consensus"

3879..4182

/note="AluU repeat: matches 1..289 of consensus"

4183..4479

/note="L1M3 repeat: matches 1903..2203 of consensus"

4480..4788

/note="AluSg repeat: matches 2..310 of consensus"

4789..5617

/note="L1M4 repeat: matches 1094..1993 of consensus"

5628..5925

/note="AluY repeat: matches 1..298 of consensus"

6403..6459

/note="MIR repeat: matches 197..259 of consensus"

6899..7218

/note="AluY repeat: matches 1..311 of consensus"

7225..7364

/note="MIR repeat: matches 46..188 of consensus"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

7961..8260

/note="AluB repeat: matches 1..300 of consensus"

8440..8736

/note="AluB repeat: matches 1..298 of consensus"

9747..9956

/note="MIR repeat: matches 36..250 of consensus"

10124..10315

/note="L1M1 repeat: matches 5988..6322 of consensus"

10384..10682

/note="AluY repeat: matches 1..300 of consensus"

10954..11235

/note="AluY repeat: matches 1..294 of consensus"

11236..11543

/note="AluSg repeat: matches 3..294 of consensus"

12286..12493

/note="MIR repeat: matches 32..249 of consensus"

13328..13622

/note="AluSg repeat: matches 1..308 of consensus"

15766..15864

/note="L2 repeat: matches 2630..2736 of consensus"

16580..16778

/note="L2 repeat: matches 2540..2742 of consensus"

17117..17212

/note="MIR repeat: matches 30..124 of consensus"

17358..18333

/note="SV4 repeat: matches 2..954 of consensus"

17886..19097

/note="CpG island"

/evidence="not-experimental"

18166..18189

/note="1372 bases of IS186 (X03123) removed here. This
 sequence represents the duplicated flanking sequence of
 the IS186."

18343..19109

/note="SV4 repeat: matches 646..1386 of consensus"

19813..21277

/note="L1M4 repeat: matches 4785..6297 of consensus"

21663..21969

/note="AluB repeat: matches 1..308 of consensus"

22425..22651

/note="L1M4 repeat: matches 4456..4697 of consensus"

23061..23181

/note="MIR repeat: matches 34..170 of consensus"

23287..23332

/note="L2 repeat: matches 2692..2737 of consensus"

23293..23345

/note="MIR repeat: matches 200..249 of consensus"

26800..27101

/note="AluSg repeat: matches 1..302 of consensus"

27275..27582

/note="AluB repeat: matches 1..306 of consensus"

27607..27900

/note="AluSg repeat: matches 1..308 of consensus"

27909..28044

/note="FLM4C repeat: matches 1..133 of consensus"

28130..28431

/note="AluB repeat: matches 6..305 of consensus"

28517..28572

/note="MIR repeat: matches 96..155 of consensus"

29159..29275

/note="AluB repeat: matches 1..118 of consensus"

29425..29729

/note="AluY repeat: matches 1..305 of consensus"

30098..30173

/note="L1M5 repeat: matches 7840..7913 of consensus"

30259..30544

/note="AluB repeat: matches 1..284 of consensus"

30739..30875

/note="L1M5 repeat: matches 7603..7731 of consensus"

30860..31103

/note="L1M4 repeat: matches 7643..7901 of consensus"

31112..31223

/note="AluB/q repeat: matches 194..305 of consensus"


```

note="Single-stranded terminator coverage."
repeat_region /rpt_family="L1MB5" 9458. .9530
repeat_region /rpt_family="AluSx" 9536. 9848
repeat_region /rpt_family="L2" 9878. .10115
repeat_region /rpt_family="L2" 10128. .10357
repeat_region /rpt_family="MER20" 10432. .10458
repeat_region /rpt_family="AT-rich" complement(10511. .10567)
repeat_region /rpt_family="AluO/FLAM" 10612. .10908
repeat_region /rpt_family="AluJb" 10913. .11210
repeat_region /rpt_family="AluSp" 11216. .11512
repeat_region /rpt_family="AluSx" 11529. .11553
repeat_region /rpt_family="(TGA)n" complement(11647. .11759)
repeat_region /rpt_family="L2" complement(11816. .12159)
repeat_region /rpt_family="L2" complement(13008. .13306)
repeat_region /rpt_family="AluSp" 13355. .13490
repeat_region /rpt_family="AluSx" 13491. .13789
repeat_region /rpt_family="AluSc" 13790. .13956
repeat_region /rpt_family="AluSx" complement(14078. .14206)
repeat_region /rpt_family="MIR" complement(14220. .14276)
repeat_region /rpt_family="L2" complement(14277. .14309)
repeat_region /rpt_family="MIR" complement(14310. .14594)
repeat_region /rpt_family="AluSp" complement(14604. .14912)
repeat_region /rpt_family="AluJo" complement(14913. .15045)
repeat_region /rpt_family="MIR" complement(15061. .15157)
repeat_region /rpt_family="MIR" complement(15517. .15637)

Query Match 64.0% Score 16; DB 9; Length 183700;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 3 gtcaagtgcaagtcag 18
|||||
Db 177274 GTCAGGTGCAGCTCAG 177259

RESULT 43
ID AC025339/c standard; DNA; HTG; 185492 BP.
XX AC025339;
XX AC025339.1
XX
XX
SV
XX
XX 14-MAR-2000 (Rel. 63, Created)
XX 14-MAR-2000 (Rel. 63, Last updated, Version 1)
DE Homo sapiens chromosome 2 clone RP11-676D22 map 2, WORKING DRAFT SEQUENCE.
XX 51 unordered pieces.
XX HTG; HTGS_DRAFT; HTGS_PHASE1.

```



```

CC * 3534 3633: gap of 100 bp
CC * 3634 4974: contig of 1341 bp in length
CC * 4975 5074: gap of 100 bp
CC * 5075 6187: contig of 1113 bp in length
CC * 6188 6287: gap of 100 bp
CC * 6288 7383: contig of 1096 bp in length
CC * 7384 7483: gap of 100 bp
CC * 7484 8508: contig of 1025 bp in length
CC * 8509 8608: gap of 100 bp
CC * 8609 9799: contig of 1191 bp in length
CC * 9800 9899: gap of 100 bp
CC * 9900 11275: contig of 1376 bp in length
CC * 11276 11375: gap of 100 bp
CC * 11376 12804: contig of 1429 bp in length
CC * 12805 12904: gap of 100 bp
CC * 12905 14246: contig of 1342 bp in length
CC * 14247 14346: gap of 100 bp
CC * 14347 15544: contig of 1198 bp in length
CC * 15545 15644: gap of 100 bp
CC * 15645 17181: contig of 1537 bp in length
CC * 17182 17281: gap of 100 bp
CC * 17282 19392: contig of 2111 bp in length
CC * 19393 19492: gap of 100 bp
CC * 19493 21710: contig of 2218 bp in length
CC * 21711 21810: gap of 100 bp
CC * 21811 24149: contig of 2339 bp in length
CC * 24150 24249: gap of 100 bp
CC * 24250 26685: contig of 2436 bp in length
CC * 26686 26785: gap of 100 bp
CC * 26786 28327: contig of 1542 bp in length
CC * 28328 28427: gap of 100 bp
CC * 28428 31535: contig of 3108 bp in length
CC * 31536 31635: gap of 100 bp
CC * 31636 34165: contig of 2530 bp in length
CC * 34166 34265: gap of 100 bp
CC * 34266 37131: contig of 2866 bp in length
CC * 37132 37231: gap of 100 bp
CC * 37232 40119: contig of 2888 bp in length
CC * 40120 40219: gap of 100 bp
CC * 40220 43234: contig of 3015 bp in length
CC * 43235 43334: gap of 100 bp
CC * 43335 46125: contig of 2791 bp in length
CC * 46126 46225: gap of 100 bp
CC * 46226 48748: contig of 2523 bp in length
CC * 48749 48848: gap of 100 bp
CC * 48849 53250: contig of 4402 bp in length
CC * 53251 53350: gap of 100 bp
CC * 53351 57172: contig of 3822 bp in length
CC * 57173 57272: gap of 100 bp
CC * 57273 60281: contig of 3009 bp in length
CC * 60282 60381: gap of 100 bp
CC * 60382 63484: contig of 3103 bp in length
CC * 63485 63584: gap of 100 bp
CC * 63585 67025: contig of 3441 bp in length
CC * 67026 67125: gap of 100 bp
CC * 67126 70475: contig of 3350 bp in length
CC * 70476 70575: gap of 100 bp
CC * 70576 74638: contig of 4063 bp in length
CC * 74639 74738: gap of 100 bp
CC * 74739 78257: contig of 3519 bp in length
CC * 78258 78357: gap of 100 bp
CC * 78358 83359: contig of 5002 bp in length
CC * 83360 83459: gap of 100 bp
CC * 83460 86986: contig of 3527 bp in length
CC * 86987 87086: gap of 100 bp
CC * 87087 90187: contig of 3101 bp in length
CC * 90188 90287: gap of 100 bp
CC * 90288 94735: contig of 4448 bp in length
CC * 94736 94835: gap of 100 bp
CC * 94836 99768: contig of 4933 bp in length
CC * 99769 99868: gap of 100 bp
CC * 99869 104278: contig of 4410 bp in length
CC * 104279 104378: gap of 100 bp

```

```

CC * 104379 110808: contig of 6430 bp in length
CC * 110809 110908: gap of 100 bp
CC * 110909 116478: contig of 5570 bp in length
CC * 116479 116578: gap of 100 bp
CC * 116579 122443: contig of 5865 bp in length
CC * 122444 122543: gap of 100 bp
CC * 122544 126331: contig of 3788 bp in length
CC * 126332 126431: gap of 100 bp
CC * 126432 131374: contig of 4943 bp in length
CC * 131375 131474: gap of 100 bp
CC * 131475 138177: contig of 6703 bp in length
CC * 138178 138277: gap of 100 bp
CC * 138278 144292: contig of 6015 bp in length
CC * 144293 144392: gap of 100 bp
CC * 144393 150325: contig of 5933 bp in length
CC * 150326 150425: gap of 100 bp
CC * 150426 156168: contig of 5743 bp in length
CC * 156169 156268: gap of 100 bp
CC * 156269 163498: contig of 7230 bp in length
CC * 163499 163598: gap of 100 bp
CC * 163599 173145: contig of 9547 bp in length
CC * 173146 173245: gap of 100 bp
CC * 173246 185492: contig of 1247 bp in length.

```

```

XX Key Location/Qualifiers
FH source 1. 185492
FT /chromosome="2"
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /map="2"
FT /clone="RP11-676D22"
FT /clone_11b="RPC1-11 Human Male BAC"
FT misc_feature 1. 11026
FT /note="assembly-fragment"
FT misc_feature 1127. 2310
FT /note="assembly-fragment"
FT misc_feature 2411. 3533
FT /note="assembly-fragment"
FT misc_feature 3634. 4974
FT /note="assembly-fragment"

```

```

Query Match 64.0%; Score 16; DB 33; Length 185492;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 caggtgcaggtcagca 20
Db 43512 CAGGTGCAGGTACGCA 43497

```

```
RESULT 44
```

```
AC064835 standard; DNA; HTG; 202828 BP.
```

```
AC064835;
```

```
AC064835.3
```

```

XX 26-APR-2000 (Rel. 63, Created)
DT 10-JUL-2000 (Rel. 64, Last updated, Version 6)
DE Homo sapiens chromosome 16 clone RP11-673P17, WORKING DRAFT SEQUENCE, 12
DE unordered pieces.
DE
XX HTG: HTGS-DRAFT; HTGS-PHASE1.
XX

```

```

OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
XX
XX [1]
RN 1-202828

```

RA Waterston R.H.:
 FT "The sequence of Homo sapiens clone";
 RL Unpublished.
 XX [2]
 RN 1-202828
 RP Waterston R.H.;
 RA Submitted (22-APR-2000) to the EMBL/GenBank/DBJ databases.
 RL Forest Sequencing Center, Washington University School of Medicine, 4444
 RL Forest Park Parkway, St. Louis, MO 63108, USA
 XX
 On May 16, 2000 this sequence version replaced gi:7712280.
 CC Genome Center
 CC Center: Washington University Genome Sequencing Center
 CC Center code: WUGSC
 CC Web site: http://genome.wustl.edu/gsc/index.shtml
 CC Project Information
 CC Center project name: H.NH0673P17
 CC ----- Summary Statistics -----
 CC Sequencing vector: M13; 100%
 CC Chemistry: Dye-Primer ET; 100% of reads
 CC Chemistry: Dye-terminator Big Dye; 0% of reads
 CC Assembly program: Phrap; version 0.990319
 CC Consensus quality: 195107 bases at least Q40
 CC Consensus quality: 197538 bases at least Q30
 CC Consensus quality: 198879 bases at least Q20
 CC Insert size: 199000; agarose-fp
 CC Insert size: 201728; sum-of-ctrls
 CC Quality coverage: 5.33 in Q20 bases; agarose-fp
 CC Quality coverage: 5.29 in Q20 bases; sum-of-ctrls
 CC -----
 CC * NOTE: This is a 'working draft' sequence. It currently
 CC * consists of 12 contigs. The true order of the pieces
 CC * is not known and their order in this sequence record is
 CC * arbitrary. Gaps between the contigs are represented as
 CC * runs of N, but the exact sizes of the gaps are unknown.
 CC * This record will be updated with the finished sequence
 CC * as soon as it is available and the accession number will
 CC * be preserved.
 CC 1
 CC 1109: contig of 1109 bp in length
 CC 1110: gap of unknown length
 CC 1210: contig of 4726 bp in length
 CC 5936: gap of unknown length
 CC 6036: contig of 6985 bp in length
 CC 13021: gap of unknown length
 CC 13121: contig of 7908 bp in length
 CC 21029: gap of unknown length
 CC 21129: contig of 6285 bp in length
 CC 27414: gap of unknown length
 CC 27514: contig of 7038 bp in length
 CC 34552: gap of unknown length
 CC 34652: contig of 18492 bp in length
 CC 53144: gap of unknown length
 CC 53244: contig of 23958 bp in length
 CC 77202: gap of unknown length
 CC 77302: contig of 28855 bp in length
 CC 106157: gap of unknown length
 CC 106257: contig of 25388 bp in length
 CC 131645: gap of unknown length
 CC 131745: contig of 26954 bp in length
 CC 158699: gap of unknown length
 CC 158799: contig of 44030 bp in length.
 XX
 FH Key Location/Qualifiers
 FH source 1. 202828
 FT /chromosome="16"
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /clone="RP11-673P17"
 FT misc_feature 1. 1109

FT misc_feature /note="assembly_name:Contig14"
 FT 1210. .5935
 FT /note="assembly_name:Contig18"
 FT 6036. .13020
 FT /note="assembly_name:Contig19"
 FT 13121. .21028
 FT /note="assembly_name:Contig20"
 FT 21129. .27413
 FT /note="assembly_name:Contig21 clone_end:77
 FT vector_side:right"
 FT 27514. .34551
 FT /note="assembly_name:Contig22"
 FT 34652. .53143
 FT /note="assembly_name:Contig23"
 FT 53244. .77201
 FT /note="assembly_name:Contig24"
 FT 77302. .106156
 FT /note="assembly_name:Contig25"
 FT 106257. .131644
 FT /note="assembly_name:Contig26 clone_end:SP6
 FT vector_side:left"
 FT 131745. .158698
 FT /note="assembly_name:Contig27"
 FT 158799. .202828
 FT /note="assembly_name:Contig28"
 XX
 SQ Sequence 202828 BP; 53011 A; 46987 C; 47342 G; 54379 T; 1109 other;

Query Match 64.0%; Score 16; DB 33; Length 202828;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 caggtgcaggtcagca 20
 Db 46988 CAGGTGCAGGTGACCA 47003
 |||||

RESULT 45
 AC016186 245499 bp DNA 12-MAR-2000
 AC016186/c Homo sapiens chromosome 18 clone RP11-8115 map 18, *** SEQUENCING
 LOCUS DEFINITION
 IN PROGRESS *** 70 unordered pieces.
 AC016186
 VERSION AC016186.3 GI:7230177
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 245499)
 Birren, B., Linton, L., Nusbaum, C., and Lander, E.
 Homo sapiens chromosome 18, clone RP11-8115
 Unpublished
 2 (bases 1 to 245499)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barina, N., Beckert, R., Boguslavsky, L., Bouckigalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
 Cooke, P., Dearellano, K., Dewar, K., Dolino, M., Donelan, L., Doyle, M.,
 Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hages, B., Heaford, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kann, L., Karatis, A., Klein, D.,
 Lebecky, J., Lieu, C., Locke, K., MacDonald, P., Margolis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C., H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 Direct Submission

TITLE
 JOURNAL Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Mar 12, 2000 this sequence version replaced gi:6649364.

All repeats were identified using RepeatMasker:
Smil, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L4258

Center clone name: 8_I_15

NOTE: This is a 'working draft' sequence. It currently consists of 70 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1174: contig of 1174 bp in length
1175 1274: gap of 100 bp
1275 2609: contig of 1335 bp in length
2610 2709: gap of 100 bp
2710 3931: contig of 1222 bp in length
3932 4031: gap of 100 bp
4032 5121: contig of 1090 bp in length
5122 5221: gap of 100 bp
5222 6287: contig of 1066 bp in length
6288 6387: gap of 100 bp
6388 7686: contig of 1299 bp in length
7687 7786: gap of 100 bp
7787 8797: contig of 1011 bp in length
8798 8897: gap of 100 bp
8898 10536: contig of 1639 bp in length
10537 10636: gap of 100 bp
10637 11858: contig of 1222 bp in length
11859 11958: gap of 100 bp
11959 13375: contig of 1417 bp in length
13376 13475: gap of 100 bp
13476 15102: contig of 1627 bp in length
15103 15202: gap of 100 bp
15203 16445: contig of 1243 bp in length
16446 16545: gap of 100 bp
16546 17637: contig of 1092 bp in length
17638 17737: gap of 100 bp
17738 19309: contig of 1572 bp in length
19310 19409: gap of 100 bp
19410 21532: contig of 2123 bp in length
21533 21632: gap of 100 bp
21633 22166: contig of 1534 bp in length
22167 23266: gap of 100 bp
23267 25262: contig of 1996 bp in length
25263 25362: gap of 100 bp
25363 27315: contig of 1953 bp in length
27316 27415: gap of 100 bp
27416 28722: contig of 1307 bp in length
28723 28822: gap of 100 bp
28823 30596: contig of 1774 bp in length
30597 30696: gap of 100 bp
30697 33203: contig of 2507 bp in length
33204 33303: gap of 100 bp
33304 35408: contig of 2105 bp in length
35409 35508: gap of 100 bp
35509 37441: contig of 1933 bp in length
37442 37541: gap of 100 bp
37542 40401: contig of 2860 bp in length
40402 40501: gap of 100 bp
40502 42052: contig of 1551 bp in length
42053 42152: gap of 100 bp
42153 44294: contig of 2142 bp in length
44295 44394: gap of 100 bp
44395 46328: contig of 1934 bp in length

46329 46428: gap of 100 bp
46429 48649: contig of 2221 bp in length
48650 48749: gap of 100 bp
48750 51037: contig of 2288 bp in length
51038 51137: gap of 100 bp
51138 53348: contig of 2211 bp in length
53349 53448: gap of 100 bp
53449 55337: contig of 1889 bp in length
55338 55437: gap of 100 bp
55438 57028: contig of 1591 bp in length
57029 57128: gap of 100 bp
57129 59830: contig of 2702 bp in length
59831 59930: gap of 100 bp
59931 62172: contig of 2242 bp in length
62173 62272: gap of 100 bp
62273 64940: contig of 2668 bp in length
64941 65040: gap of 100 bp
65041 67862: contig of 2822 bp in length
67863 67962: gap of 100 bp
67963 70236: contig of 2274 bp in length
70237 70336: gap of 100 bp
70337 72533: contig of 2197 bp in length
72534 72633: gap of 100 bp
72634 74796: contig of 2163 bp in length
74797 74896: gap of 100 bp
74897 77899: contig of 3003 bp in length
77900 77999: gap of 100 bp
78000 81075: contig of 3076 bp in length
81076 81175: gap of 100 bp
81176 84346: contig of 3171 bp in length
84347 84446: gap of 100 bp
84447 87026: contig of 2580 bp in length
87027 87126: gap of 100 bp
87127 90859: contig of 3733 bp in length
90860 90959: gap of 100 bp
90960 93599: contig of 2640 bp in length
93600 93699: gap of 100 bp
93700 96946: contig of 3247 bp in length
96947 97046: gap of 100 bp
97047 100076: contig of 3030 bp in length
100077 100176: gap of 100 bp
100177 105333: contig of 5157 bp in length
105334 105433: gap of 100 bp
105434 109169: contig of 3736 bp in length
109170 109269: gap of 100 bp
109270 113940: contig of 4671 bp in length
113841 114040: gap of 100 bp
114041 119295: contig of 5255 bp in length
119296 119395: gap of 100 bp
119396 120878: contig of 1483 bp in length
120879 120978: gap of 100 bp
120979 123579: contig of 2601 bp in length
123580 123679: gap of 100 bp
123680 128211: contig of 4532 bp in length
128212 128311: gap of 100 bp
128312 132536: contig of 4225 bp in length
132537 132636: gap of 100 bp
132637 137226: contig of 4590 bp in length
137227 137326: gap of 100 bp
137327 142521: contig of 5195 bp in length
142522 142621: gap of 100 bp
142622 147070: contig of 4449 bp in length
147071 147170: gap of 100 bp
147171 151866: contig of 4656 bp in length
151867 151966: gap of 100 bp
151967 156914: contig of 4948 bp in length
156915 157014: gap of 100 bp
157015 162849: contig of 5835 bp in length
162850 162949: gap of 100 bp
162950 169655: contig of 6706 bp in length
169656 169755: gap of 100 bp
169756 174040: contig of 4285 bp in length
174041 174140: gap of 100 bp

```
* 174141 180979: contig of 6839 bp in length
* 180980 181079: gap of 100 bp
* 181080 190751: contig of 9672 bp in length
* 190752 190851: gap of 100 bp
* 190852 202044: contig of 11193 bp in length
* 202045 202144: gap of 100 bp
* 202145 211327: contig of 9183 bp in length
* 211328 211427: gap of 100 bp
* 211428 224152: contig of 12725 bp in length
* 224153 224252: gap of 100 bp
* 224253 234336: contig of 9984 bp in length
* 234257 234337: gap of 100 bp
* 234337 245499: contig of 11163 bp in length.
```

FEATURES

```
source
1..245499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-8115"
/clone_1bp="RPC1-11 Human Male BAC"
```

```
Query Match 64.0%; Score 16; DB 2; Length 245499;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 5 caagtgcaqtlcaaca 20
|||||
Db 109551 CAGGTGCAGGTACCA 109536
```

Search completed: December 26, 2001, 12:54:24
Job time: 8986 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:17:08 : Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196f-10

Perfect score: 1 tgcgcagtcgcagtcgcagtcgctg 25

Scoring table: OLIGO_NUC
Gapop 60.0, Capext 60.0

Searched: 930621 seqs, 428662619 residues

Word size: 9

Total number of hits satisfying chosen parameters: 47864

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database:

N.Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AAAX01303 E. coli biotin syn
3	25	100.0	1084	10	AAAN91329 E. coli B10 B gene.
4	25	100.0	1121	7	AAAN60496 Sequence encoding
5	25	100.0	5872	15	AAO62386 Biotin-biosynthesis
6	16	64.0	450	22	AA112043 Probe #1976 for ge
7	16	64.0	450	22	AA133378 Probe #2064 used t
8	16	64.0	450	22	AA101968 Probe #1959 used t
9	15	60.0	333	21	AAO56641 Eucalyptus grandis
10	15	60.0	630	16	AAO78948 Human immunoglobul
11	15	60.0	1969	22	AA114105 Probe #4038 for ge

C	12	15	60.0	1969	22	AA135486
C	13	15	60.0	1969	22	AA103958
C	14	15	60.0	4674	20	AAAX90421
C	15	14	56.0	269	21	AAAB31747
C	16	14	56.0	327	21	AAAC00048
C	17	14	56.0	363	22	AAH42390
C	18	14	56.0	374	22	AAAF64981
C	19	14	56.0	423	14	AAO45597
C	20	14	56.0	423	20	AAZ39427
C	21	14	56.0	423	20	AAZ23965
C	22	14	56.0	444	18	AA180181
C	23	14	56.0	447	20	AAZ20406
C	24	14	56.0	613	16	AAO78943
C	25	14	56.0	981	21	AAAI5905
C	26	14	56.0	1100	21	AAAI5915
C	27	14	56.0	1130	21	AAAB39064
C	28	14	56.0	1135	21	AAAB39087
C	29	14	56.0	1161	21	AAAB91019
C	30	14	56.0	1173	21	AAZ98136
C	31	14	56.0	1186	21	AAAC98064
C	32	14	56.0	1186	22	AAH35018
C	33	14	56.0	1191	19	AAAT98726
C	34	14	56.0	1368	22	AAE44623
C	35	14	56.0	1368	22	AAZ96242
C	36	14	56.0	1634	19	AAV37360
C	37	14	56.0	1814	22	AA159259
C	38	14	56.0	1816	22	AA161045
C	39	14	56.0	37856	21	AAAI1992
C	40	14	56.0	534720	19	AAV30458
C	41	14	56.0	536165	14	AAV30459
C	42	13	52.0	18	12	AAO13647
C	43	13	52.0	123	20	AAAB8925
C	44	13	52.0	153	22	AA123772
C	45	13	52.0	153	22	AA149080
C	46	13	52.0	153	22	AA109377
C	47	13	52.0	225	14	AAO60177
C	48	13	52.0	269	22	AA122207
C	49	13	52.0	276	21	AAAC00049
C	50	13	52.0	279	22	AA126888
C	51	13	52.0	279	22	AA15684
C	52	13	52.0	363	20	AAAX9527
C	53	13	52.0	363	20	AAAX9527
C	54	13	52.0	370	22	AAH34737
C	55	13	52.0	375	21	AAAP22004
C	56	13	52.0	378	18	AAAT61237
C	57	13	52.0	378	18	AAAT61239
C	58	13	52.0	386	22	AA114576
C	59	13	52.0	386	22	AA135944
C	60	13	52.0	386	22	AA104386
C	61	13	52.0	395	22	AA112969
C	62	13	52.0	400	21	AAAC00033
C	63	13	52.0	423	20	AAAX79207
C	64	13	52.0	437	21	AAZ42889
C	65	13	52.0	463	22	AA116085
C	66	13	52.0	463	22	AA118303
C	67	13	52.0	472	21	AAAT5419
C	68	13	52.0	477	20	AAZ20408
C	69	13	52.0	480	22	AAH65984
C	70	13	52.0	527	20	AAV87894
C	71	13	52.0	529	22	AAH13160
C	72	13	52.0	546	16	AAO78964
C	73	13	52.0	553	22	AAH97949
C	74	13	52.0	588	22	AA117685
C	75	13	52.0	588	22	AA142619
C	76	13	52.0	593	22	AA140100
C	77	13	52.0	600	22	AAH66012
C	78	13	52.0	699	21	AAAO2409
C	79	13	52.0	718	22	AAAB9146
C	80	13	52.0	723	22	AAAF70988
C	81	13	52.0	732	22	AAH70228
C	82	13	52.0	750	20	AAAX58277
C	83	13	52.0	879	16	AAAT01341
C	84	13	52.0	880	21	AAE20929

Probe #4172 used t
Probe #3949 used t
Human ataxin-2 lik
Plant microsatelli
Human secreted pro
Nucleotide sequenc
Novel human polynu
Sequence encoding
Antibody B17X2 VH
Human B17X2 antibo
Monoclonal antibod
IgG antibody 2.4.4
Human immunoglobul
Human protein clon
Human secreted pro
Human secreted pro
Human signal pepti
Human colon cancer
Human colon cancer
DNA encoding a S.
Novel protein kina
S. pneumoniae deri
Streptococcus pneu
Human polynucleoti
S. cellulosum DNA
Rhizobium species
Rhizobium species
Exon B/intron junc
Oligonucleotide us
Probe #13705 for g
Probe #17766 used
Probe #9368 used t
Human brain expres
Probe #12140 for g
Human secreted pro
Probe #16821 for g
Probe #24370 used
DNA encoding the h
Heavy chain coding
Human colon cancer
Human breast and o
Human anti-RSV mon
Human anti-RSV mon
Probe #4509 for ge
Probe #4630 used t
Probe #4377 used t
Probe #2902 for ge
Human secreted pro
Anti-HIV-1 gp120 a
Human 5' EST isola
Probe #6018 for ge
Probe #6989 used t
Human OREF ORP974
IgG antibody 2.3.2
C glutamicum codin
EST clone FG380.
Human immunoglobul
Murine 7-transmemb
Probe #7618 for ge
Probe #11305 used
Probe #8786 used t
C glutamicum codin
Human colon cancer
Polyketide synthas
C. glutamicum SKT
Human cervical can
Zea mays ScIBr pro
Human FK506 bindin
Human FK-506 bindi

```

C 85 13 52.0 880 21 AAA4807 Human adenosine re
C 86 13 52.0 911 22 AAH29728 S cerevisiae apopl
C 87 13 52.0 933 22 AAH32118 Human olfactory re
C 88 13 52.0 939 22 AAH67432 C glutamicum codin
C 89 13 52.0 952 22 AAH32030 Human olfactory re
C 90 13 52.0 963 22 AAH67430 C glutamicum codin
C 91 13 52.0 969 22 AAF28571 Fragment of pyclic
C 92 13 52.0 1040 20 AAH58275 Zea mays SCLBt pro
C 93 13 52.0 1101 19 AAH34298 Human secreted pro
C 94 13 52.0 1113 22 AAF71611 Corynebacterium 91
C 95 13 52.0 1121 22 AAD07933 Human TDC protein-
C 96 13 52.0 1122 21 AAZ98342 A. thaliana gene 1
C 97 13 52.0 1160 21 AAZ93322 Glycine max iso1a
C 98 13 52.0 1173 22 AAC08663 Human DIB2 cDNA.
C 99 13 52.0 1215 22 AAH45073 Murine FM-3 coding
C 100 13 52.0 1282 22 AAF28572 Fragment of pyclic

```

ALIGNMENTS

```

RESULT 1
ID AAI62941 standard; DNA; 839 BP.
AC AAI62941;
DT 22-OCT-2001 (first entry)
DE Human genomic DNA SEQ ID NO 269.
XX
KW Human; nootropic; neuroprotective; cytosolic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antischizoid; antianemic; antiallergic; cancer;
KW antihemetic; hepatotoxic; cerebroprotective; antidiabetic; cancer;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotoxic; gene therapy; vaccine;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200155449-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01346.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0218290.
PR 14-AUG-2000; 2000US-0225447.
PR 01-SEP-2000; 2000US-0229343.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 25-SEP-2000; 2000US-0234997.
PR 29-SEP-2000; 2000US-0236367.
PR 13-OCT-2000; 2000US-0239937.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249265.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.

```

```

PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
PI WPI, 2001-476225/51.
XX
DR Novel plasma membrane associated proteins useful for diagnosing,
XX treating, preventing and/or prognosing disorders related to the
XX proteins, including cancer, immune response and neuronal disorders -
XX
PS Example 2: SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAI62752-AAI62961) and proteins
XX (AAH2347-AAH42415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral ischaemia;
CC and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other:

```

```

Query Match 100.0%; Score 25; DB 22; Length 839;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 tgcgcaggtgcagtcagcagcttg 25
Db 143 tgcgcaggtgcagtcagcagcttg 167

```

```

RESULT 2
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
XX
PR 08-MAR-1995; 95US-0401068.
XX

```

PA	(NOVS) NOVARTIS FINANCE CORP.
XX	
PI	Patton DA:
XX	
DR	WPI: 1999-152902/13.
DR	P-PSDB: AAW73906.
XX	
PT	Transgenic plants with high biotin levels - transformed with DNA
PT	encoding di:amino-pelargonic acid amino-transferase or biotin
PT	synthase
XX	
PS	Example 2: Column 37-40; 34pp: English.
XX	
CC	This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC	be used in the transgenic plant of the invention. The transgenic plant,
CC	plant cell or plant tissue is transformed with a chimeric gene encoding
CC	di:amino-pelargonic acid (DAP) aminotransferase or biotin synthase and
CC	produces more biotin than a non-transgenic plant, cell or tissue. The
CC	plant is used as an improved dietary source of biotin (vitamin H) for
CC	humans or animals.
XX	
SO	Sequence 1041 BP: 262 A; 273 C; 305 G; 201 T; 0 other:
QY	
QY	1 tcgtcagtgccaggtcagcagcttg 25
Db	111 tcgtcagtgccaggtcagcagcttg 135
RESULT	3
AA	AAAN91329
ID	AAAN91329 standard: DNA; 1084 BP.
XX	
AC	AAAN91329:
XX	
DT	15-FEB-1990 (first entry)
XX	
DE	E.coli Bio B gene.
XX	
KW	E.coli; Bio B gene; biotin.
XX	
OS	Escherichia coli.
XX	
PH	Key Location/Qualifiers
FT	CDS 24..1064
FT	/tag=a
XX	
PN	GB2216530-A.
XX	
PD	11-OCT-1989.
XX	
PF	17-MAR-1989; 89GB-0006210.
XX	
PR	22-MAR-1988; 88GB-0006804.
XX	
PR	17-MAR-1989; 89GB-0006210.
XX	
PA	(UKAG-) UK MIN. AGRIC. FISH.
XX	
PI	Pearson BM, McKee RA:
XX	
DR	WPI: 1989-295085/41. P-PSDB P91392
XX	
PT	Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT	- derived from E.coli and capable of replication and expression in other
PT	microorganisms, esp. yeast.
XX	
PS	Table 3; page 33-4; 52pp: English.
XX	
CC	The gene can be used in a plasmid for expression of enzymes of the biotin

```

CC syntheticlipthway. Pref. control sequences for expression in S.cerevisiae
CC CC plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC CC Lactobacillus. Insertion of bio B improves biotin yields in
CC CC microorganisms which export biotin, or enables growth in media contg.
CC CC little or no biotin of organisms unable to synthesise biotin for their
XX CC own use.
SO Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match          100.0%; Score 25; DB 10; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 tcgtcaggtcagtcagtcagcagcttg 25
         |||
Db       134 tcgtcaggtcagtcagtcagcagcttg 158

RESULT 4
AAN60496 AAN60496 standard; DNA; 1121 BP.
XX
AC AAN60496;
XX
DT 17-OCT-1991 (first entry)
XX
DE Sequence encoding biotin synthesising enzyme.
XX
KW Biotin synthetlc enzyme; E.coli; desthiobiotin; ds.
XX
FH Key Location/Qualifiers
FT CDS 42..1082
FT /*tag= a
XX
PN JP61149091-A.
XX
PD 07-JUL-1986.
XX
PF 24-DEC-1984; 84JP-0272605.
XX
PR 24-DEC-1984; 84JP-0272605.
XX
PA (NIPS ) NIPPON SODA KK.
XX
WI 1986-216622/33.
XX
DR P-PSDB: AAP60535.
XX
PT Double stranded DNA encoding biotin synthesising enzyme -
PT PT comprises transformed mutant E.coli strain contg. cyclic doubled
PT PT stranded DNA encoding biotin synthetlc biotin enzyme as a plasmid.
XX
PS
XX
PS Disclosure: Page 534; 23pp; Japanese.
XX
CC The sequence may be expressed by a transformed E.coli host, cultured
CC CC in a medium containing desthiobiotin.
XX
SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match          100.0%; Score 25; DB 7; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 tcgtcagtcagtcagtcagcagcttg 25
         |||
Db       152 tcgtcagtcagtcagtcagcagcttg 176

RESULT 5
AAN62386 AAN62386 standard; DNA; 5872 BP.
XX
```

AC	AA062386;	
XX		
DT	16-NOV-1994	(first entry)
XX		
DE	Biotin-biosynthesis genes conty. plasmid pB030A-15/9.	
XX		
XX	Biotin; expression; enterobacteriæ; vitamin H; synthesis;	
KW	plasmid; pB030A-15/9; biob; bioF; bioc; biod; bioa;	
KW	promoter plac; biotin synthase; KAPA synthase;	
KW	8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;	
KW	dethiobiotin synthase; DAPA synthase;	
KW	S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;	
KW	seborrhoea; dermatitis; ds.	
XX		
OS	Escherichia coli DSM498.	
FH		
XX		
Key	location/Qualifiers	
FT	1..96	
FT	/*tag= a	
FT	/function= "promoter plac"	
FT	/evidence= EXPERIMENTAL	
FT	23..28	
FT	/*tag= b	
FT	/standard_name= "promoter plac"	
FT	45..50	
FT	/*tag= c	
FT	/evidence= EXPERIMENTAL	
FT	/standard_name= "promoter plac"	
FT	105..109	
FT	/*tag= d	
FT	/evidence= EXPERIMENTAL	
FT	/standard_name= "biob RBS no. 9"	
FT	117..1157	
FT	/*tag= e	
FT	/product= "biotin synthase"	
FT	/evidence= EXPERIMENTAL	
FT	/gene= "biob"	
FT	/number= 1	
FT	1141..1146	
FT	/*tag= f	
FT	/standard_name= "biob RBS"	
FT	1154..2311	
FT	/*tag= g	
FT	/EC_number= 2.3.1.47	
FT	/product= "KAPA synthase"	
FT	/evidence= EXPERIMENTAL	
FT	/gene= "biob"	
FT	/number= 2	
FT	/standard_name= "8-amino-7-oxononanoate synthase"	
FT	2284..2288	
FT	/*tag= h	
FT	/standard_name= "bioc RBS"	
FT	2295..3050	
FT	/*tag= i	
FT	/function= "involved in pimeloyl-CoA synthesis"	
FT	/product= "protein"	
FT	/gene= "bioc"	
FT	/number= 3	
FT	3030..3033	
FT	/*tag= j	
FT	/standard_name= "biob RBS"	
FT	3043..3753	
FT	/*tag= k	
FT	/EC_number= 6.3.3.3	
FT	/product= "DTB synthase"	
FT	/evidence= EXPERIMENTAL	
FT	/gene= "biob15"	
FT	/number= 4	
FT	/standard_name= "dethiobiotin synthase"	
FT	3712..3750	
FT	/*tag= l	
FT	/note= "biob15 substitution"	
FT	3742..3746	
RBS		

FT	/tag= m
FT	/standard_name= "bioA RBS"
FT	3750..5039
CDS	
FT	/tag= n
FT	/EC_number= 2.6.1.62
FT	/product= "DAPA synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "bioA"
FT	/number= 5
FT	/standard_name= "S-adenosyl-L-methionine: 8-amino-
FT	-7-oxononanoate aminotransferase"
RBS	
FT	5088..5093
FT	/tag= o
FT	/standard_name= "ORF1 RBS"
FT	5098..5574
CDS	
FT	/tag= p
FT	/function= "unknown, involved in biotin synthesis"
FT	/product= "protein"
FT	/evidence= EXPERIMENTAL
FT	/gene= "ORF1"
FT	/number= 6
FT	5583..5644
FT	/tag= q
FT	/standard_name= "rho-independent transcriptional
FT	terminator"
stem_loop	
FT	5583..5605
FT	/tag= r
PN	
PD	WO9408023-A.
PD	14-APR-1994.
Pf	01-OCT-1993; 93WO-EP02688.
PR	02-OCT-1992; 92CH-0003124.
PR	15-JUL-1993; 93CH-0002134.
PA	(LONZ) LONZA AG.
PI	Birch O, Brass J, Fuhrmann M, Shaw N;
DR	WPI: 1994-135587/16.
DR	P-FSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
PT	Biotechnological biotin prodn. using enterobacterial biotin-gene
PT	- providing vitamin H in high yield
PS	Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
CC	The sequence is derived from plasmid pB030A-15/9 contg. the
CC	bioB, bioF, bioC, bioD and bioE genes responsible for biosynthesis
CC	of biotin, arranged in a transcription unit. Microorganisms
CC	contg. these DNA fragments or plasmids may be used in the prodn.
CC	of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC	loss of appetite and tiredness.
SQ	Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other:
Query Match	100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity	100.0%; Pred. No. 0.00032;
Matches 25; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Oy	1 tgcgcaggtcagtcacgacttg 25
I	
Db	227 tgcgcaggtcagtcacgacttg 251
RESULT	6
AAI12043/c	
ID	AAI12043 standard; DNA: 450 BP.
XX	
XC	AAI12043.


```

XX 12-OCT-2001 (first entry)
DT
XX
XX Probe #1976 for gene expression analysis in human cervical cell sample.
DE
XX Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
OS
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 1976; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENP). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at fcp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 450 BP; 121 A; 117 C; 124 G; 88 T; 0 other;
XX
XX
XX Query Match 64.0%; Score 16; DB 22; Length 450;
XX Best Local Similarity 100.0%; Pred. No. 12;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 3 gtcagtgacgtcag 18
XX |||||||
XX Db 423 GTCAGTGACGTCAG 408
XX
XX
XX RESULT 7
XX AA13378/c
XX ID AA13378 standard; DNA: 450 BP.
XX
XX AA13378;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #2064 used to measure gene expression in human placenta sample.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX

```

```

PN WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 2064; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 450 BP; 121 A; 117 C; 124 G; 88 T; 0 other;
XX
XX
XX Query Match 64.0%; Score 16; DB 22; Length 450;
XX Best Local Similarity 100.0%; Pred. No. 12;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 3 gtcagtgacgtcag 18
XX |||||||
XX Db 423 GTCAGTGACGTCAG 408
XX
XX
XX RESULT 8
XX AA101968/c
XX ID AA101968 standard; DNA: 450 BP.
XX
XX AA101968;
XX
XX 09-OCT-2001 (first entry)
XX
XX Probe #1959 used to measure gene expression in human breast sample.
XX
XX Probe: human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX

```

XX Penn SG, Hanzel DK, Chen W, Rank DK;
 XX WPI; 2001-476286/51.
 XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 25; SEQ ID No 1959; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridizes at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pot_sequences.
 CC
 SQ Sequence 450 BP; 121 A; 117 C; 124 G; 88 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 450;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gtcaggtgcaggtcag 18
 |||
 DB 423 gtcaggtgcaggtcag 408

RESULT 9
 AAC56641/c
 ID AAC56641 standard; DNA; 353 BP.
 AC AAC56641;
 AC
 DT 25-JAN-2001 (first entry)
 XX
 DE Eucalyptus grandis transcription factor DNA sequence #512.
 XX
 KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MIB; SS.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO200053724-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 09-MAR-2000; 2000WO-US06112.
 XX
 PR 11-MAR-1999; 99US-026513.
 XX
 PR 18-AUG-1999; 99US-0149485.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX WPI; 2000-579369/54.
 XX
 PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide

PT
 XX
 PS Claim 1; Page 482; 747pp; English.
 XX
 CC The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC binding families of regulatory proteins: bZIP, bZIP family of G-box
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MIB.
 CC
 SQ Sequence 353 BP; 117 A; 59 C; 98 G; 78 T; 1 other;

Query Match 60.0%; Score 15; DB 21; Length 353;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcgtcaggtcaggt 15
 |||
 DB 326 TCGTCAGGTGCGAGT 312

RESULT 10
 AA078948/c
 ID AA078948 standard; DNA; 630 BP.
 AC AA078948;
 AC
 DT 01-AUG-1995 (first entry)
 XX
 DE Human immunoglobulin Vh gene #10.
 XX
 KW Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
 KW cosmid; placenta; vector; pUB81; E.coli; mammalian; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 71..495
 FT CDS
 FT
 FT /*tag= a
 FT /product= human immunoglobulin variable heavy chain
 FT 114..199
 FT /*tag= b
 FT 322..324
 FT /*tag= c
 FT /transl_except= unused termination codon
 FT 373..375
 FT /*tag= d
 FT /transl_except= unused termination codon
 FT 496..498
 FT /*tag= e
 FT /note= "miscellaneous signal, does not conform to
 FT termination or splice site sequence"

PN WO9426895-A.
 XX
 PD 24-NOV-1994.
 XX
 PF 10-MAY-1993; 93WO-JP00603.
 XX
 PR 10-MAY-1993; 93WO-JP00603.
 XX
 PA (NISB) JAPAN TOBACCO INC.
 XX
 PI Honjo T, Matsuda F;
 XX

DR WPI: 1995-006791/01.
 DR P-PSDB: AAR66304.
 XX DNA fragment comprising human immunoglobulin Vh genes - for the
 PT production of human immunoglobulin in mammalian hosts
 XX
 PS Claim 19: Page 43-44; 130pp; Japanese.
 CC A series of genes (AA078939-79002) encoding human immunoglobulin
 CC variable heavy chains. The genes were isolated and cloned from a series
 CC of cosmid constructs: Y202; Y103; Y21; Y5; Y24; 3-31; M84; M118 and M131,
 CC by PCR amplification using primers AA078917-38. The genes are subdivided
 CC into 5 families of Vh genes. The fragments cover a region of 800 kb.
 CC The DNA fragments were isolated from high molecular weight DNA from
 CC human placenta. The DNA was partially digested with Taqi restriction
 CC enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
 CC fractions were collected. The fragments were ligated with ClaI-digested
 CC cosmid vector pJB81. The ligation products were then subcloned by colony
 CC infected into E.coli 490A. The fragments were then subcloned by colony
 CC hybridisation. The Vh genes and the DNA fragments encoding them are
 CC useful in producing human immunoglobulin in mammalian hosts.
 XX
 SQ Sequence 630 BP; 141 A; 179 C; 148 G; 162 T; 0 other;
 Query Match 60.0%; Score 15; DB 16; Length 630;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 6 agtgcagtcagcga 20
 DB 280 AGGTGAGGTGAGCA 266
 |||||||||||||||
 RESULT 11
 AA114105/C
 ID AA114105 standard: DNA; 1969 BP.
 XX
 AC AA114105;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Probe #4038 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 XX
 PR 26-MAY-2000; 2000US-0207456.
 XX
 PR 30-JUN-2000; 2000US-0608408.
 XX
 PR 03-AUG-2000; 2000US-0632366.
 XX
 PR 21-SEP-2000; 2000US-0234687.
 XX
 PR 27-SEP-2000; 2000US-0236359.
 XX
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI: 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 25; SEQ ID NO 4038; 487pp; English.

XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human Hela cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1969 BP; 551 A; 432 C; 640 G; 346 T; 0 other;
 Query Match 60.0%; Score 15; DB 22; Length 1969;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 9 tgcaggtcagcagct 23
 DB 1418 TGCAGGTGAGCAGCT 1404
 |||||||||||||||
 RESULT 12
 AA135486/C
 ID AA135486 standard: DNA; 1969 BP.
 XX
 AC AA135486;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #4172 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 XX
 PR 26-MAY-2000; 2000US-0207456.
 XX
 PR 30-JUN-2000; 2000US-0608408.
 XX
 PR 03-AUG-2000; 2000US-0632366.
 XX
 PR 21-SEP-2000; 2000US-0234687.
 XX
 PR 27-SEP-2000; 2000US-0236359.
 XX
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI: 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID NO 4172; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENPs).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 1969 BP; 551 A; 432 C; 640 G; 346 T; 0 other;

```

Query Match          60.0%; Score 15; DB 22; Length 1969;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 tgcaggtcagcagc 23
    |||||
Db 1418 TGCAGGTGACGACGT 1404

RESULT 13
AA103958/c
ID AA103958 standard; DNA: 1969 BP.
XX
XX AA103958;
AC
XX 09-OCT-2001 (first entry)
DE
XX Probe #3949 used to measure gene expression in human breast sample.
XX
XX Probe: human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
OS
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DK;
XX
XX WPI: 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast.
XX
XX Claim 25; SEQ ID No 3949; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1969 BP; 551 A; 432 C; 640 G; 346 T; 0 other;

Query Match          60.0%; Score 15; DB 22; Length 1969;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 tgcaggtcagcagc 23
    |||||
Db 1418 TGCAGGTGACGACGT 1404

```

```

RESULT 14
AAx90421/c
ID AAx90421 standard; cDNA: 4674 BP.
XX
XX AAx90421;
AC
XX 29-SEP-1999 (first entry)
DE
XX Human ataxin-2 like gene.
XX
XX Human; ataxin-2 like gene; ataxin-2 like protein; ataxia; deafness;
XX cardiomyopathy; neurological disease; cancer; AIDS; diagnosis; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 63..3218
XX /*tag= a
XX /product= "ataxin-2 like protein"

WO936527-A1.
XX
XX 22-JUL-1999.
XX
XX 19-JAN-1998; 98WO-CN00009.
XX
XX 19-JAN-1998; 98WO-CN00009.
XX
XX (UYHU-) UNIV HUNAN MEDICAL.
XX
XX Deng H, Liu C, Wang D, Xia J;
XX
XX WPI: 1999-458463/38.
XX
XX P-PSDB; AAY29321.
XX
XX Ataxin-2 like protein, and related polynucleotides, useful in
XX treatment and diagnosis of ataxia,
XX
XX Claim 4; Page 21-23; 34pp; English.
XX
XX The present sequence represents a human ataxin-2 like gene. Ataxin-2
XX like polynucleotides (I), vectors containing (I) and recombinant host
XX cells are useful for recombinant production of ataxin-2 like protein.
XX Both (I), ataxin-2 like protein and antibodies against ataxin-2 like
XX protein are useful as research reagents, for screening assays and in
XX diagnostic assays. Antagonists and agonists of ataxin-2 like protein
XX can be used to inhibit or enhance, respectively, the activity of
XX ataxin-2 like protein or expression of (I). Anti-ataxin-2 like protein
XX antibodies and ataxin-2 like protein or its fragments can be used
XX in vaccines. In particular, the proteins, antibodies, agonists and
XX antagonists can be used for treating, e.g. ataxia, cardiomyopathy,
XX deafness, neurological disease, cancer and AIDS, related to both an
XX excess and insufficient amounts of ataxin-2 like.
XX
XX Sequence 4674 BP; 971 A; 1566 C; 1244 G; 893 T; 0 other;

Query Match          60.0%; Score 15; DB 20; Length 4674;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caggtgacgagcagc 19
    |||||
Db 3741 CAGGTGACGAGTACG 3727

RESULT 15
AAA31747
ID AAA31747 standard; DNA: 269 BP.
XX
XX AAA31747;
AC

```

```

XX 05-JUL-2000 (first entry)
XX
XX Plant microsatellite marker #708.
DE
XX Plant microsatellite sequence: core repeat sequence; detection; probe;
XX DNA polymorphism; genome mapping; physical mapping; fingerprinting;
XX variety identification; genetic variability evaluation; primer; ss.
XX
XX Eucalyptus grandis.
OS
XX W09967421-A1.
XX
XX 29-DEC-1999.
XX
XX 25-JUN-1999; 99WO-NZ00092.
XX
XX 25-JUN-1998; 98US-0105307.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Havukkala IJ, Bloksberg LN, Glenn M:
XX WPI, 2000-116958/10.
XX
XX New plant microsatellite markers and associated flanking species for
XX the detection of polymorphic genetic markers -
XX
XX Claim 1: Page 285; 392pp; English.
XX
XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
XX and associated flanking species. The sequences comprise a central core
XX repeat sequence, especially selected from the sequences AAA32094-A32096
XX with left and right flanking sequences. The polynucleotide sequences
XX can be used in the detection of DNA polymorphisms, in genome mapping,
XX in physical mapping, in positional cloning of genes, in variety
XX identification and in evaluation of genetic variability within and
XX between plant tissues, populations, cultivars, species and species
XX groups. They may also be used to design hybridization probes for
XX oligonucleotide fingerprinting and library screening and to design
XX primers for microsatellite-primed PCR. Microsatellite markers are
XX useful to locate specific economically useful genes in plant genomes.
XX
XX Sequence 269 BP; 63 A; 62 C; 61 G; 83 T; 0 other:
SQ

```

Query Match 56.0%; Score 14; DB 21; Length 269;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 11 caggtcagcagctt 24
   |||||
Db 76 caggtcagcagctt 89

```

RESULT 16
 AAC00048/C
 ID AAC00048 standard; cDNA: 327 BP.
 XX
 AC AAC00048:
 XX
 XX 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 46.
 XX
 KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 XX gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 XX EPI033401-A2.
 XX
 XX

```

PD 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
XX
XX P-PSDB: AAG00042.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1: SEQ ID 46; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. An ORF has been identified within the
XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
XX derived from 30 different tissues. EST sequences usually correspond
XX mainly to the 3' untranslated region (UTR) of the mRNA because they are
XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
XX well suited for isolating cDNA sequences derived from the 5' ends of
XX mRNAs and even in those cases where longer cDNA sequences have been
XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
XX cDNAs with intact 5' ends and can therefore be used to obtain full length
XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
XX gene therapy and chromosome mapping procedures. They are used to obtain
XX upstream regulatory sequences and to design expression and secretion
XX vectors.
XX
XX Sequence 327 BP; 70 A; 102 C; 75 G; 69 T; 11 other:
SQ

```

Query Match 56.0%; Score 14; DB 21; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 6 aggtgcagctcagc 19
   |||||
Db 144 AGCTGCAGCTCAGC 131

```

RESULT 17
 AAH42390/C
 ID AAH42390 standard; DNA: 363 BP.
 XX
 AC AAH42390:
 XX
 XX 01-OCT-2001 (first entry)
 XX
 DE Nucleotide sequence of variable heavy chain fragment of clone G81.
 XX
 KM Antibody; heavy chain; VH; amyloid protein; blood brain barrier;
 XX endometrial cell; brain cell antigen; inflammation; adhesion molecule;
 KM transferrin receptor; neurological disease; Alzheimer's disease;
 XX prion disease; AIDS-related dementia; epilepsy; brain injury; ss.
 XX
 OS Homo sapiens.
 XX
 XX W0200144300-A2.
 XX
 XX 21-JUN-2001.
 XX
 XX 27-NOV-2000; 2000WO-GB04501.
 XX
 XX 13-DEC-1999; 99US-0170599.
 XX
 XX (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
 XX
 XX

PI Webster C, Osbourn J, Ward G, Miller K;
 XX
 DR WPI: 2001-396131/42.
 XX P-PSDB: AAG62946.
 PT Mixture or panel of antibodies for selecting specific binding members
 PT that cross the blood brain barrier, for use in delivering different
 PT molecules and treating neurological diseases
 PS
 XX
 PS Disclosure: Page 98; 109pp; English.

CC The present sequence encodes an antibody variable heavy chain (VH)
 CC fragment. The fragment is used to produce a mixture or panel of 5
 CC different specific binding members, each comprising an antibody VH
 CC and/or VL variable domain and capable, when displayed on the surface
 CC of filamentous bacteriophage particles or in the case of a specific
 CC binding member comprising the D5 VH and/or VL variable domain when
 CC bound to human serum amyloid protein, to pass through a mammalian
 CC blood brain barrier (BBB). The panel is useful for the selection of
 CC specific binding members with a desired property such as ability to
 CC cross BBB, ability to bind endothelial cells or other brain cell antigen,
 CC ability to bind areas of inflammation in the brain or BBB breakdown or
 CC ability to bind intracellular adhesion molecules and to bind transthyrin
 CC receptor. The antibodies are useful in diagnosis, prophylaxis and
 CC treatment of human or animal body, including neurological diseases, such
 CC as Alzheimer's disease, prion disease, AIDS-related dementia, epilepsy
 CC and traumatic brain injury and any diseases involving inflammation
 CC occurring within the brain or central nervous system.
 SQ Sequence 363 BP; 84 A; 108 C; 96 G; 75 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 363;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcagtcagc 19
 |||||
 DB 70 AGGTGCAGTCAGC 57

RESULT 18
 AAF64981/c
 ID AAF64981 standard; cDNA: 374 BP.
 XX
 AC AAF64981;

XX 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 737.

KW Human: cytostatic; gene therapy: colon cancer; prostate cancer;
 KW breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PN WO200102568-A2.

XX 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.

XX 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinget J, Kassam A;
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Dmanac R;
 PI Crnenjakov R, Dmanac S, Dickson M, Labat I, Leshkowitz D;
 PI Kita D, Garcia V, Jones LW, Strache-Crain B;

DR WPI: 2001-091805/10.

XX
 PT Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -
 PS
 XX Claim 9; Page 650; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

SQ Sequence 374 BP; 71 A; 116 C; 118 G; 68 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 374;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcagtcagc 19
 |||||
 DB 180 AGGTGCAGTCAGC 167

RESULT 19
 AAQ45597/c
 ID AAQ45597 standard; DNA: 423 BP.
 XX
 AC AAQ45597;

XX 04-DEC-1993 (first entry)

DE Sequence encoding the VH of antibody B17X2.

KW Variable heavy antibody chain; human subgroup 4 germline; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 7..423

FT /*tag= a

PN WO9312231-A.

XX 24-JUN-1993.

PF 13-DEC-1991; 91WO-AU00583.

XX 13-DEC-1991; 91WO-AU00583.

PA (DOWC) DOW CHEM AUSTRALIA LTD.

PI Johnson KS, Mezes PS, Richard RA;

XX WPI: 1993-214173/26.

DR P-PSDB: AAR38315.

XX New composite antibody binding to tumor associated TAG-72
 PT antigen - includes light chain variable region from human
 PT subgroup 4 germline gene, useful, opt. as conjugate, for
 PT diagnosis or treatment of cancer

```

XX PS Disclosure; Figure 4: 150bp; English.
XX CC Cell line B17X2 expresses an antibody utilizing a variable light
CC chain encoded by a gene derived from Humd VL and a variable heavy
CC chain which makes a stable VL and VH combination.
XX CC
SQ Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other:

Query Match
Best Local Similarity 56.0%; Score 14; DB 14; Length 423;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 aggtgcaggtcagc 19
    |||
Db 133 AGGTGCAGGTCAAGC 120

RESULT 20
AAZ39427/C
ID AAZ39427 standard; DNA: 423 BP.
XX AC
XX AC AAZ39427;
XX AC
XX AC 29-FEB-2000 (first entry)
XX AC
XX DE Antibody B17X2 VH nucleotide sequence.
XX DE
XX DE
XX DE Tumor associated sialylated glycoprotein; TAG-72; cancer antigen;
XX KW carcinoma lesion; diagnostic; cancer; antibody; human; B17X2;
XX KW anti-mouse antibody hypersensitivity reaction; ss.
XX KW
XX OS Homo sapiens.
XX OS
XX PN US5976845-A.
XX PD 02-NOV-1999.
XX PD
XX PF 07-JUN-1995; 95US-0487743.
XX PF
XX PR 16-JUN-1994; 94US-0261354.
XX PR 19-APR-1990; 90US-0510697.
XX PR 20-OCT-1992; 92US-0964536.
XX PR
XX PA (DOWC ) DOW CHEM CO.
XX PA
XX PI Johnson KS, Richard RA, Mezes PS;
XX PI WPI; 1999-619651/53.
XX DR P-PSDB; AAY57179.
XX DR
XX PT Production of humanized anti-TAG-72 antibodies, used for the detection,
XX PT in vivo imaging and treatment of cancers -
XX PS Disclosure; Fig 4A-B; 85pp; English.
XX PS
XX CC The invention relates to producing humanized anti-tumor associated
XX CC sialylated glycoprotein (TAG-72) antibodies (anti-TAG). The antibodies
XX CC have binding specificity for the cancer antigen TAG-72. These antibodies
XX CC have variable regions with VL segments derived from human subgroup IV
XX CC germ-line gene and a VH segment (encoded by the VH1phatAG germ-line gene)
XX CC which is capable of combining with the VL to form a three dimensional
XX CC structure having the ability to bind TAG-72. They can be used for the in
XX CC vivo detection of carcinoma lesions. They can also be used for in vitro
XX CC diagnostics. They can also be modified with therapeutic agents e.g. a
XX CC radionuclide, drug, biological response modifier, toxin or another
XX CC antibody for the treatment of cancers. The humanized anti-TAG-72
XX CC antibodies can reduce harmful anti-mouse antibody hypersensitivity
XX CC reactions.
XX CC
SQ Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other:

```

```

XX PS Query Match
XX PS Best Local Similarity 56.0%; Score 14; DB 20; Length 423;
XX PS Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 aggtgcaggtcagc 19
    |||
Db 133 AGGTGCAGGTCAAGC 120

RESULT 21
AAZ23965/C
ID AAZ23965 standard; DNA: 423 BP.
XX AC
XX AC AAZ23965;
XX AC
XX AC 09-FEB-2000 (first entry)
XX AC
XX DE Human B17X2 antibody VH segment DNA.
XX DE
XX DE
XX DE Human; antibody; humanized; anti-tumor; sialylated glycoprotein antibody;
XX KW TAG-72; cytostatic; cancer antigen; detection; carcinoma lesion;
XX KW diagnostic; treatment; ss.
XX KW
XX OS Homo sapiens.
XX OS
XX PN US5976531-A.
XX PD 02-NOV-1999.
XX PD
XX PF 16-JUN-1994; 94US-0261354.
XX PF
XX PR 19-APR-1990; 90US-0510697.
XX PR 20-OCT-1992; 92US-0964536.
XX PR
XX PA (DOWC ) DOW CHEM CO.
XX PA
XX PI Johnson KS, Mezes PS, Richard RA;
XX PI WPI; 1999-632731/54.
XX DR P-PSDB; AAY50688.
XX DR
XX PT New humanized anti-TAG-72 antibodies, used for the detection, in vivo
XX PT imaging and treatment of cancers -
XX PS Disclosure; Figure 4A-B; 83pp; English.
XX PS
XX CC This invention describes novel humanized anti-tumor associated
XX CC sialylated glycoprotein antibodies (TAG-72) which have cytostatic
XX CC activity. The antibodies have binding specificity for the cancer antigen
XX CC TAG-72. They can be used for the in vivo detection of carcinoma lesions.
XX CC They can also be used for in vitro diagnostics. They can also be modified
XX CC with therapeutic agents e.g. a radionuclide, drug, biological response
XX CC modifier, toxin or another antibody for the treatment of cancers. The
XX CC humanized anti-TAG-72 antibodies can reduce harmful anti-mouse antibody
XX CC hypersensitivity reactions.
XX CC
SQ Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other:

Query Match
Best Local Similarity 56.0%; Score 14; DB 20; Length 423;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 aggtgcaggtcagc 19
    |||
Db 133 AGGTGCAGGTCAAGC 120

RESULT 22
AAT80181/C
ID AAT80181 standard; DNA: 444 BP.
XX AC

```


XX Primer: PCR; amplify: human immunoglobulin; variable; heavy chain;
 KW cosmid; placenta; vector: pJB81; E.coli; mammalian; ds.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 71..513
 FT /tag= a
 FT /product= human immunoglobulin variable heavy chain
 FT intron 117..202
 FT /tag= b
 FT misc_signal 514..516
 FT /tag= c
 FT /note= "miscellaneous signal, does not conform to
 FT terminator or splice site sequence"
 FT
 PN W09426895-A.
 PD 24-NOV-1994.
 XX
 XX 10-MAY-1993; 93WO-JP00603.
 XX
 XX 10-MAY-1993; 93WO-JP00603.
 XX
 PA (NISB) JAPAN TOBACCO INC.
 XX
 PI Honjo T, Matsuda F;
 DR WPI: 1995-006791/01.
 DR P-PSDB: AAR66299.
 XX
 PT DNA fragment comprising human immunoglobulin Vh genes - for the
 PT production of human immunoglobulin in mammalian hosts
 XX
 PS Disclosure: Page 36-37; 130pp: Japanese.
 XX
 PS A series of genes (AA078939-79002) encoding human immunoglobulin
 CC variable heavy chains. The genes were isolated and cloned from a series
 CC of cosmid constructs: Y202; Y103; Y21; Y6; Y24; 3-31; M84; M18 and M31.
 CC by PCR amplification using primers AA078917-38. The genes are subdivided
 CC into 5 families of Vh genes. The fragments cover a region of 800 kb.
 CC The DNA fragments were isolated from high molecular weight DNA from
 CC human placenta. The DNA was partially digested with TaqI restriction
 CC enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
 CC fragments were collected. The fragments were ligated with ClaI-digested
 CC cosmid vector pJB81. The ligation products were in vitro packed and
 CC injected into E.coli 490A. The fragments were then subcloned by colony
 CC hybridisation. The Vh genes and the DNA fragments encoding them are
 CC useful in producing human immunoglobulin in mammalian hosts.
 CC
 CC Sequence 613 BP: 139 A; 194 C; 139 G; 141 T; 0 other:
 SO
 Query Match 56.0%; Score 14; DB 16; Length 613;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 6 aggtcagtcagtcagc 19
 |||||||||||||
 Db 283 AGGTCTCAGTCAGTCAGC 270
 RESULT 25
 AAA15905/c
 ID AAA15905 standard; cDNA: 981 BP.
 AC
 XX AAA15905;
 XX
 XX 12-JUN-2000 (first entry)
 XX
 DE Human protein clone HP10568 coding sequence.
 XX

KW Human protein; hydrophobic domain; nutritional source; haematopoiesis;
 KW cytokine production; cell proliferation; cell differentiation;
 KW immune deficiency; infectious disease; autoimmune disorder; asthma;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
 KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
 KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
 KW coagulation disorder; myocardial infarction; inflammatory condition;
 KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
 KW nephritis; therapy; ss.
 XX
 XX Homo sapiens.
 OS
 XX W0200005367-A2.
 PN
 PD 03-FEB-2000.
 XX
 XX 22-JUL-1999; 99WO-JP03929.
 PF
 XX 24-JUL-1998; 98JP-0208820.
 PR 07-AUG-1998; 98JP-0224105.
 PR 25-AUG-1998; 98JP-0238116.
 PR 09-SEP-1998; 98JP-0254736.
 PR 29-SEP-1998; 98JP-0275505.
 XX
 PA (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.
 XX
 PI Kato S, Kimura T;
 DR WPI: 2000-182694/16.
 DR P-PSDB: AAY94857.
 XX
 PT Novel human proteins having hydrophobic domains useful for treating
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
 XX
 PS Claim 3; Page 189; 351pp: English.
 XX
 CC This sequence encodes a human protein of the invention, which has
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat
 CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of haematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC tooth repair processes. The protein is used in the treatment of nervous
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and
 CC Huntington's disease. They are useful for protection or regeneration and
 CC treatment of lung or liver fibrosis, reperfusion injury in various
 CC tissues, and conditions resulting from systemic cytokine damage. They are
 CC also used for promoting or inhibiting tissue differentiation. They are
 CC also used as contraceptives since they exhibit activin or inhibin related
 CC activities and as a fertility inducing therapeutic. They are used for
 CC treating various coagulation disorders and in treatment and prevention of
 CC conditions resulting from coagulation activities e.g. myocardial
 CC infarction or stroke. They also acts as receptors, receptor ligands or
 CC inhibitors or agonists of receptor/ligand interactions. They are used to
 CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
 CC reperfusion injury, arthritis, and nephritis. They can be used to

CC prevent tumours.
 XX Sequence 981 BP; 194 A; 299 C; 275 G; 213 T; 0 other;
 SQ

Query Match 56.0%; Score 14; DB 21; Length 981;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtcagtcacg 19
 |||||
 DB 139 AGCTGCAGTCACG 126

RESULT 26
 AAA15915/c
 ID AAA15915 standard; cDNA: 1100 BP.
 XX
 AC AAA15915;
 XX
 DT 12-JUN-2000 (first entry)
 XX
 DE Human protein clone Hp10568 full length coding sequence.
 XX
 KW Human protein; hydrophobic domain; nutritional source; haematopoiesis;
 KW cytokine production; cell proliferation; cell differentiation;
 KW immune deficiency; infectious disease; autoimmune disorder; asthma;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
 KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
 KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
 KW coagulation disorder; myocardial infarction; inflammatory condition;
 KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
 KW nephritis; therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200005367-A2.
 XX
 PD 03-FEB-2000.
 XX
 PE 22-JUL-1999; 99WO-JP03929.
 XX
 PF 24-JUL-1998; 98JP-0208820.
 PR 07-AUG-1998; 98JP-0224105.
 PR 25-AUG-1998; 98JP-0238116.
 PR 09-SEP-1998; 98JP-0254736.
 PR 29-SEP-1998; 98JP-0275505.
 XX
 FA (SAGA) SAGAMI CHEM RES CENT.
 FA (PROT-) PROTEGENE INC.
 XX
 PI Kato S, Kimura T;
 XX
 DR WPI: 2000-182694/16.
 DR P-PSDB: AAY94857.
 XX
 PT Novel human proteins having hydrophobic domains useful for treating
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
 XX
 PS Claim 4; Page 203-205; 351pp; English.

This sequence encodes a human protein of the invention, which has
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat

CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of haematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC tooth repair processes. The protein is used in the treatment of nervous
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and
 CC Huntington's disease. They are useful for protection or regeneration and
 CC treatment of lung or liver fibrosis, reperfusion injury in various
 CC tissues, and conditions resulting from systemic cytokine damage. They are
 CC also used for promoting or inhibiting tissue differentiation. They are
 CC also used as contraceptives since they exhibit activin or inhibin related
 CC activities and as a fertility inducing therapeutic. They are used for
 CC treating various coagulation disorders and in treatment and prevention of
 CC conditions resulting from coagulation activities e.g. myocardial
 CC infarction or stroke. They also acts as receptors, receptor ligands or
 CC inhibitors or agonists of receptor/ligand interactions. They are used to
 CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
 CC reperfusion injury, arthritis, and nephritis. They can be used to
 CC prevent tumours.
 XX
 SQ Sequence 1100 BP; 221 A; 333 C; 314 G; 232 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1100;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtcagtcacg 19
 |||||
 DB 195 AGCTGCAGTCACG 182

RESULT 27
 AAA39064/c
 ID AAA39064 standard; cDNA: 1130 BP.
 XX
 AC AAA39064;
 XX
 DT 30-AUG-2000 (first entry)
 XX
 DE Human secreted protein gene 13 SEQ ID NO:23.
 XX
 KW Human; secreted protein; cytosolic; anti-proliferative; vulnary;
 KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
 KW hyperproliferative disorder; infectious disease; tissue regeneration;
 KW hyper-vascular disease; chromosome 11; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200017222-A1.
 XX
 PD 30-MAR-2000.
 XX
 PE 22-SEP-1999; 99WO-US22012.
 XX
 PF 23-SEP-1998; 98US-0101546.
 PR 02-OCT-1998; 98US-0102895.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
 XX Komatsoulis G, Endress GA, Soppet DR;
 XX WPI: 2000-283538/24.
 XX P-PSDB: AAB08903.

PT Human secreted proteins and coding sequences useful in diagnostic and
 PT therapeutic methods for disorders such as immune system or
 PT proliferative disorders, related to the proteins
 PS Claim 1: Page 329-330; 416pp; English.
 XX
 CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the
 CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytostatic;
 CC anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The
 CC secreted proteins and their related polynucleotide sequences are useful
 CC for diagnostic and therapeutic methods useful for diagnosing and treating
 CC disorders related to the secreted proteins. The proteins, and
 CC polynucleotide sequences may be useful for treating disorders of the
 CC immune system, hyperproliferative disorders, infectious disease,
 CC regeneration of tissues, for chemotaxis and for screening molecules that
 CC bind to the proteins. The proteins or polynucleotide sequences may be
 CC used as food additives or preservatives, to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, co-factors or other nutritional components. Agonists or
 CC antagonists of the proteins may be used to prevent scar tissue growth
 CC during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
 CC and AAB08890 are sequences used in the exemplification of the present
 CC invention.
 CC
 XX
 SQ Sequence 1130 BP; 246 A; 334 C; 317 G; 233 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1130;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcagc 19
 |||
 DB 201 AGGTGCAGGTCAAGC 188

RESULT 28
 AAA39087/C
 ID AAA39087 standard; cDNA: 1135 BP.
 XX
 AC AAA39087;
 XX
 DT 30-AUG-2000 (first entry)
 XX
 DE Human secreted protein gene 13 SEQ ID NO:46.
 XX
 KW Human; secreted protein; cytostatic; anti-proliferative; vulnerary;
 KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
 KW hyperproliferative disorder; infectious disease; tissue regeneration;
 KW screening; food additive; preservative; wound healing;
 KW hyper-vascular disease; chromosome 11; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200017222-A1.
 XX
 PD 30-MAR-2000.
 XX
 PE 22-SEP-1999; 99WO-US22012.
 XX
 PR 23-SEP-1998; 98US-0101546.
 PR 02-OCT-1998; 98US-0102895.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
 PI Komatsoulis G, Endress GA, Soppet DR.
 XX
 DR WPI; 2000-283538/24.
 DR P-PSDB; AAB08926.
 XX

PT Human secreted proteins and coding sequences useful in diagnostic and
 PT therapeutic methods for disorders such as immune system or
 PT proliferative disorders, related to the proteins
 PS Claim 1: Page 347-348; 416pp; English.
 XX

CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the
 CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytostatic;
 CC anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The
 CC secreted proteins and their related polynucleotide sequences are useful
 CC for diagnostic and therapeutic methods useful for diagnosing and treating
 CC disorders related to the secreted proteins. The proteins, and
 CC polynucleotide sequences may be useful for treating disorders of the
 CC immune system, hyperproliferative disorders, infectious disease,
 CC regeneration of tissues, for chemotaxis and for screening molecules that
 CC bind to the proteins. The proteins or polynucleotide sequences may be
 CC used as food additives or preservatives, to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, co-factors or other nutritional components. Agonists or
 CC antagonists of the proteins may be used to prevent scar tissue growth
 CC during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
 CC and AAB08890 are sequences used in the exemplification of the present
 CC invention.
 CC
 XX

SQ Sequence 1135 BP; 243 A; 333 C; 323 G; 234 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 1135;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcagc 19
 |||
 DB 188 AGGTGCAGGTCAAGC 175

RESULT 29
 AAA91019/C
 ID AAA91019 standard; DNA: 1161 BP.
 XX
 AC AAA91019;
 XX
 DT 05-APR-2001 (first entry)
 XX
 DE Human secreted protein PRO7154 coding sequence.
 XX
 KW Secreted protein; human; PRO protein; neoplastic cell growth; tumour;
 KW proliferation; leukaemia; lymphoid malignancy; inflammatory disorder;
 KW angiogenic disorder; immunologic disorder; PRO7154; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT CDS Location/Qualifiers
 FT 102..1085
 FT /*tag= a
 FT /product= PRO7154
 PN W0200075317-A2.
 XX
 PD 14-DEC-2000.
 XX
 PE 15-MAY-2000; 2000WO-US13358.
 XX
 PR 09-JUN-1999; 99US-0138385.
 PR 20-JUL-1999; 99US-0144790.
 PR 03-AUG-1999; 99US-0146843.
 PR 10-AUG-1999; 99US-0148188.
 PR 17-AUG-1999; 99US-0149320.
 PR 17-AUG-1999; 99US-0149327.
 PR 17-AUG-1999; 99US-0149396.
 PR 20-AUG-1999; 99US-0150114.

PR 31-AUG-1999; 99US-0151700.
 PR 31-AUG-1999; 99US-0151734.
 XX
 XX (GENE) GENENTECH INC.
 PA
 PI Borstein DA, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WL;
 XX WPI: 2001-071075/08.
 DR P-PSDB; AAY97585.
 XX
 PT Antibodies against PRO polypeptides, useful for diagnosing and treating
 PT tumours are associated with gene amplification, neoplastic cell growth
 PT and proliferation in mammals -
 XX
 PS Claim 50; Fig 11; 143pp; English.
 XX
 CC This sequence encodes a human PRO protein of the invention. The PRO
 CC proteins are secreted proteins. Antagonists or antibodies of PRO
 CC polypeptides are useful for diagnosing and treating tumours are
 CC associated with gene amplification, neoplastic cell growth and
 CC proliferation in mammals, and those conditions characterised by
 CC overexpression and/or activation of the amplified genes. Such conditions
 CC include benign or malignant tumours (e.g. renal, liver, kidney, bladder,
 CC breast, gastric, ovarian, colorectal, prostate, pancreatic, lung, vulval,
 CC thyroid, hepatic carcinomas, sarcomas, glioblastomas and various head and
 CC neck tumours); leukaemias and lymphoid malignancies; neuronal, glial,
 CC astrocytic, hypothalamic, and other glandular, macrophageal, epithelial,
 CC stromal and blastocytic disorders; and inflammatory, angiogenic and
 CC immunologic disorders. These may further be used to qualitatively or
 CC quantitatively detect the expression of proteins encoded by the
 CC amplified genes, and in tumour diagnostics or prognostics. The PRO
 CC polypeptide or its antagonist may be used for the preparation of a
 CC medicament in the treatment of a condition, which is responsive to the
 CC PRO polypeptide, its antagonist or anti-PRO antibody.
 XX
 S0 Sequence 1161 BP; 235 A; 356 C; 327 G; 242 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 1161;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcagc 19
 |||||
 DB 240 AGGTGCAGGTCAAGC 227

RESULT 30
 AA298136/C
 ID AA298136 standard; cDNA; 1173 BP.
 XX
 AC AA298136;
 XX
 DT 11-MAY-2000 (first entry)
 XX
 DE Human signal peptide containing protein HSP-28 cDNA SEQ ID NO:162.
 XX
 KW Human; signal peptide-containing protein; HSP; diagnosis; cancer;
 KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
 KW antimicrobial; neurotropic; neuroprotective; cardiovascular; hepatotropic;
 KW antidiabetic; gene therapy; cell proliferation; neurological disorder;
 KW reproductive disorder; developmental disorder; arteriosclerosis;
 KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
 KW Parkinson's disease; Huntington's disease; ovulatory defect;
 KW muscular dystrophy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200000610-A2.
 XX
 PD 06-JAN-2000.
 XX

PF 25-JUN-1999; 99WO-US14484.
 XX
 XX 26-JUN-1998; 98US-0090762.
 PR 31-JUL-1998; 98US-0094983.
 PR 01-OCT-1998; 98US-0102686.
 PR 11-DEC-1998; 98US-0112129.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
 PI Akedilom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JU;
 PI Bandman O;
 XX WPI: 2000-160673/14.
 DR P-PSDB; AAY87251.
 XX
 PT New human signal peptide-containing proteins useful in treatment,
 PT prevention and diagnosis of e.g. cancer, inflammation and
 PT cardiovascular disease -
 XX
 PS Claim 9; Page 268; 327pp; English.
 XX
 CC AA298109 to AA298242 encode AAY87224 to AAY87357 which represent the
 CC human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have
 CC anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,
 CC neuroprotective, cardiovascular and antidiabetic activities, and can
 CC be used in gene therapy. HSPs can be used to treat or prevent disorders
 CC associated with decreased activity or function of HSP. Antagonists of
 CC HSP are used to treat or prevent disorders associated with increased
 CC activity or function of HSP. Such diseases include cell proliferation
 CC (including cancer), inflammation, cardiovascular, neurological,
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
 CC asthma, Crohn's disease, microbial or other infections, congestive or
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP
 CC nucleic acids can be used for the recombinant production of HSP, for
 CC detecting HSP in standard hybridisation and amplification assays (for
 CC diagnosis and monitoring), in gene therapy, as antisense,
 CC triplex-forming or ribozyme therapeutics, for detecting related sequences
 CC or genetic variations, and for chromosomal mapping. HSP are also used to
 CC raise specific antibodies (Ab) and to screen for agonists and
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
 CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic
 CC antagonists, in competitive drug screens, and for purification of HSP
 CC from natural sources.
 XX
 S0 Sequence 1173 BP; 236 A; 363 C; 330 G; 244 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1173;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcagc 19
 |||||
 DB 256 AGGTGCAGGTCAAGC 243

RESULT 31
 AAC98064/C
 ID AAC98064 standard; cDNA; 1186 BP.
 XX
 AC AAC98064;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human colon cancer antigen nucleotide sequence SEQ ID NO:74.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioprotective; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; anti-infective; antibacterial; gene therapy; wound;

KM neural disorder; immune system disorder; muscular disorder;
 KM reproductive disorder; gastrointestinal disorder; renal disorder;
 KM infectious disease; cardiovascular disorder; ss.
 XX Homo sapiens.
 XX MO200055351-A1.
 XX PD 21-SEP-2000.
 XX PF 08-MAR-2000; 2000WO-US05883.
 XX PR 12-MAR-1999; 99US-0124270.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Ruben SM;
 XX DR WPI: 2000-587534/55.
 XX DR P-PSDB: AAB53307.
 PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 PS Claim 1: Page 520-521; 2104pp; English.
 XX AC AAG97991 to AAC98763 encode the human colon cancer associated proteins,
 XX CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 XX CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 XX CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 XX CC and vulnerability, nephrotoxic, anti-infective and antibacterial activities, and
 XX CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 XX CC proteins and antibodies to the proteins are useful for the prevention,
 XX CC treatment and diagnosis of colon disorders, such as colon cancer. The
 XX CC polynucleotides may be used in diagnostics and research, such as for
 XX CC chromosome identification, and as hybridisation probes. The proteins
 XX CC may also be used to prevent diseases such as neural disorders, immune
 XX CC system disorders, muscular disorders, reproductive disorders,
 XX CC gastrointestinal disorders, wounds, renal disorders, infectious
 XX CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 XX CC AAB54007 represent sequences used in the exemplification of the present
 XX CC invention.
 XX SQ Sequence 1186 BP; 244 A; 361 C; 338 G; 242 T; 1 other;
 Query Match 56.0%; Score 14; DB 21; Length 1186;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 aggtgcaggtcagc 19
 Db 255 AGCTGCAGGTCAAC 242
 RESULT 32
 AAH35018/C
 ID AAH35018 standard; cDNA: 1186 BP.
 XX AC AAH35018;
 XX XX
 XX DT 03-SEP-2001 (first entry)
 XX DE Human colon cancer antigen encoding cDNA SEQ ID NO:2100.
 XX XX
 XX KM Human; colon cancer; colon cancer antigen; diagnosis; detection;
 XX KM colorectal carcinoma; chromosome 11; ss.
 XX OS Homo sapiens.
 XX FT
 XX PN MO200122920-A2.
 XX XX

PD 05-APR-2001.
 XX XX
 XX PF 28-SEP-2000; 2000WO-US26524.
 XX XX
 XX PR 29-SEP-1999; 99US-0157137.
 XX PR 03-NOV-1999; 99US-0163280.
 XX XX
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX PT WPI: 2001-235357/24.
 XX DR P-PSDB: AAG75613.
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 PS Claim 1: Page 3586; 9803pp; English.
 XX AC AAG32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 XX CC the proteins are collectively known as colon cancer antigens. The colon
 XX CC cancer antigens have cytostatic activity and can be used in gene
 XX CC therapy and vaccine production. N and P may be used in the prevention,
 XX CC diagnosis and treatment of diseases associated with inappropriate P
 XX CC expression. For example, N and P may be used to treat disorders
 XX CC associated with decreased expression by rectifying mutations or deletions
 XX CC in a patient's genome that affect the activity of P by expressing
 XX CC inactive proteins or to supplement the patient's own production of P.
 XX CC Additionally, N may be used to produce the colon cancer-associated P,
 XX CC by inserting the nucleic acids into a host cell and culturing the cell
 XX CC to express the proteins. N and P can be used in the prevention, diagnosis
 XX CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 XX CC and AAB77789 represent sequences used in the exemplification of the
 XX CC present invention.
 XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 XX CC missing at time of publication, meaning no sequences are present for
 XX CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX SQ Sequence 1186 BP; 244 A; 362 C; 338 G; 242 T; 0 other;
 Query Match 56.0%; Score 14; DB 22; Length 1186;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 aggtgcaggtcagc 19
 Db 255 AGCTGCAGGTCAAC 242
 RESULT 33
 AAT98726
 ID AAT98726 standard; DNA: 1191 BP.
 XX AC AAT98726;
 XX XX
 XX DT 09-NOV-1998 (first entry)
 XX DE DNA encoding a S. pneumoniae protein of unknown function.
 XX XX
 XX KM Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 XX KM immunological response; inoculation; antibody production; inhibitor;
 XX KM T cell immune response; antimicrobial compound; bacterial adhesion;
 XX KM extracellular matrix protein; protein-mediated cell invasion; wound;
 XX KM pathogenesis; ss.
 XX KM
 XX OS Streptococcus pneumoniae.
 XX XX
 XX FH Key Location/Qualifiers
 XX FT 596..979
 XX PN /*tag= a
 XX XX

PN WO9743303-A1.
 XX 20-NOV-1997.
 PD 14-MAY-1997; 97WO-US07950.
 PF 14-MAY-1996; 96US-0017670.
 PR 14-MAY-1996; 96US-0017670.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.;
 XX
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Stodola RK;
 DR WPI: 1998-008793/01.
 DR P-PSDB: AAW38682.
 XX
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 XX
 PS Claim 4; Pages 222-223; 483pp; English.
 XX
 CC This sequence encodes a Streptococcus pneumoniae protein of unknown
 CC function, and represents a DNA sequence of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 CC
 SO Sequence 1191 BP; 294 A; 258 C; 294 G; 345 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 1191;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgtcaggtcaggt 15
 |||||||
 DB 954 cgtcaggtcaggt 967

RESULT 34
 AAF44623/C
 ID AAF44623 standard; cDNA; 1368 BP.
 XX
 AC AAF44623;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Novel protein kinase cDNA, SEQ ID NO: 2.
 XX
 KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
 KW immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
 KW dermatological; antidiabetic; antinefritility; gene therapy; vaccine;
 KW immune disorder; cardiovascular disease; neurodegenerative disease;
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
 XX
 OS Mus musculus.
 XX

PN WO200073469-A2.
 XX 07-DEC-2000.
 PD 26-MAY-2000; 2000WO-US14842.
 PF 28-MAY-1999; 99US-0136503.
 PR (SUGEN-) SUGEN INC.
 XX
 PA PLOWMAN GD, Martinez R, Whyte D, Sudersanam S;
 PI WPI: 2001-032161/04.
 DR P-PSDB: AAB65598.
 XX
 PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
 PT treating immune-related diseases and disorders, cardiovascular disease,
 PT neurodegenerative diseases and/or cancers -
 XX
 PS Disclosure; Fig 2; 310pp; English.
 XX
 CC The present sequence encodes a novel protein kinase. The nucleic acids
 CC and the protein kinases they encode may be used in the treatment and
 CC diagnosis of diseases associated with inappropriate kinase expression
 CC such as immune-related diseases and disorders, cardiovascular disease,
 CC neurodegenerative diseases and/or cancers. The nucleic acids and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays. The kinase polypeptides may be used as antigens in the production
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies
 CC and kinase antagonists may also be used to down regulate kinase
 CC expression and activity. Diseases related to kinase expression and
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
 CC disorders, complications of organ transplantation, myocardial infarction,
 CC immune disorders, cardiomyopathies, strokes, renal failure,
 CC oxidative-stress related disorders, chronic inflammatory bowel disease,
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
 CC reproductive disorders.
 CC
 SO Sequence 1368 BP; 346 A; 359 C; 398 G; 265 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1368;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtcaggtcagc 19
 |||||||
 DB 1022 AGGTGACAGTCAGC 1009

RESULT 35
 AA296242
 ID AA296242 standard; DNA; 1634 BP.
 XX
 AC AA296242;
 XX
 DT 10-APR-2000 (first entry)
 XX
 DE S. pneumoniae derived DNA from ORF #70.
 XX
 KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
 KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9806734-A1.
 XX
 PD 19-FEB-1998.
 PF 15-AUG-1997; 97WO-US14436.
 PR 16-AUG-1996; 96US-0024022.

XX (SMK) SMITHKLINE BEECHAM CORP.
 PA Black MT., Hodgson JE., Knowles DJC, Lonetto MA, Nicholas RO;
 PI Strodia RK;
 XX WPI: 1998-159452/14.
 DR P-PSDB: AAY85867.
 XX Streptococcus pneumoniae proteins and related DNA - useful for
 PT screening compounds for antibacterial activity
 XX Claim 4; Page 101-102; 640pp; English.
 PS This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see AAZ96173-296494) and their encoded proteins (see
 CC AAY85792-Y86182). The DNA, vectors and host cells described in the
 CC method of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.
 XX Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;
 SO

Query Match 56.0%; Score 14; DB 19; Length 1634;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 cgtcaggtgcaggt 15
 |||||||||||||
 DB 954 cgtcaggtgcaggt 967

RESULT 36
 AAV37360
 ID AAV37360 standard; DNA: 1634 BP.
 XX AAV37360;
 AC 13-OCT-1998 (first entry)
 XX Streptococcus pneumoniae coding region.
 DE Streptococcus pneumoniae coding region.
 KW coding region; ORF; open reading frame; antibacterial;
 KW infection; prevention; meningitis; ss.
 XX Streptococcus pneumoniae.
 OS
 XX key Location/Qualifiers
 FH 1040..1291
 FT /*tag= a
 XX /product= unknown
 XX MO9819689-A1.
 PD 14-MAY-1998.
 XX 27-OCT-1997; 97WO-US19226.
 PF 01-NOV-1996; 96US-0029930.
 XX (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX Black MT., Hodgson JE., Knowles DJC, Lonetto MA, Nicholas RO;
 PI

PI Reid RH, Zarfos PN;
 XX WPI: 1998-286586/25.
 DR P-PSDB: AAW60963.
 XX New isolated nucleic acids from Streptococcus pneumoniae - useful,
 PT e.g. for identifying anti-bacterial(s) for treatment and prevention
 of meningitis
 XX Claim 1; Page 60; 130pp; English.
 PS The sequence is that of a coding region isolated from
 CC S. pneumoniae. Its encoded protein, or agonists of it,
 CC may be useful as an antibacterial for treatment or
 CC prevention of infection, specifically caused by S. pneumoniae
 CC (particularly meningitis) but possibly also Helicobacter
 CC pylori (ulcers and gastric cancer). It may be of particular
 CC use before insertion of an in-dwelling device or any other
 CC invasive procedure. The protein, or nucleic acid encoding
 CC it, can also be used in vaccines to induce a cellular
 CC and/or humoral immune response, or to screen for other
 CC antibacterials. The DNA may also contain flanking sequences
 CC that are potential sources of control elements for bacterial
 CC gene expression. Detecting a sequence encoding the protein
 CC can be used diagnostically, e.g. to detect a mutation for
 CC serotyping or classifying infectious agents.
 XX Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;
 SO

Query Match 56.0%; Score 14; DB 19; Length 1634;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 cgtcaggtgcaggt 15
 |||||||||||||
 DB 954 cgtcaggtgcaggt 967

RESULT 37
 AA159259/G
 ID AA159259 standard; cDNA: 1814 BP.
 XX AA159259;
 AC 22-OCT-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 1462.
 DE Human polynucleotide SEQ ID NO 1462.
 XX Human; nucleotide; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW Leukemia; ss.
 XX Homo sapiens.
 OS
 XX WO200153312-A1.
 PN 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US34263.
 PF 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX

PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB: AAM40103.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1: SEQ ID NO 1462; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 1814 BP; 447 A; 531 C; 481 G; 355 T; 0 other;
 XX
 QY Query Match 56.0%; Score 14; DB 22; Length 1814;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 213 AGGTGCGAGTGCAGC 200
 XX
 RESULT 38
 ID AA161045 standard; cDNA; 1816 BP.
 AC
 AC AA161045;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 5034.
 XX
 KW Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-0534263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB: AAM41889.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1: SEQ ID NO 5034; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 1816 BP; 355 A; 481 C; 531 G; 449 T; 0 other;
 XX
 QY Query Match 56.0%; Score 14; DB 22; Length 1816;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1604 aggtgcaggtcagc 1617
 XX
 RESULT 39
 ID AA11992 standard; DNA; 37856 BP.
 AC
 AC AA11992;
 XX
 DT 07-AUG-2000 (first entry)
 XX
 DE S. cellulosum DNA encoding polyketide and heteropolyketide enzymes.
 XX
 KW Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;
 KW epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;
 KW plant-protection; ds.
 XX
 OS Sorangium cellulosum.
 XX
 FH Key Location/Qualifiers
 FH CDS complement (3398..6100)
 FT /*tag= a
 FT /product= "ORF1-trna synthetase"
 FT /note= "gtg start codon"
 FT complement (6374..7111)
 FT /*tag= b
 FT /product= "ORF2-monooxygenase"
 FT complement (8433..9550)
 FT /*tag= c


```

FT      /product= "ORF3-aminotransferase"
FT      /note= "AGT start codon given in the specification"
FT      CDS
FT      /tag= d
FT      /product= "ORF4- tyrosine/DOPA-Decarboxylase"
FT      /note= "GTG start codon"
FT      12212..13658
FT      /tag= e
FT      /product= "ORF5-3-oxoacyl-ACP-reductase"
FT      /note= "ACC start codon"
FT      15374..19984
FT      /tag= f
FT      /product= "ORF6-polypeptide synthase"
FT      20003..27889
FT      CDS
FT      /tag= g
FT      /product= "ORF7-peptide synthetase"
FT      28251..29400
FT      /tag= h
FT      /product= "ORF8-transpeptidase"
FT      complement (30040..31720)
FT      CDS
FT      /tag= i
FT      /product= "ORF9-regulation element"
FT      /note= "CGC stop codon"
FT      31982..32932
FT      /tag= j
FT      /product= "ORF10-transcription regulator"
FT      33128..33613
FT      /tag= k
FT      /product= "ORF11-regulation element"
FT      /note= "GTG start codon"
FT      33661..34077
FT      /tag= l
FT      /product= "ORF12-regulation element"
FT      complement (35255..35616)
FT      CDS
FT      /tag= m
FT      /product= "ORF13-transcription regulator"
FT      complement (35730..36242)
FT      /tag= n
FT      /product= "ORF14-transcription regulator"
FT      /note= "GTG start codon"
FT      CDS
FT      DE19846493-A1.
FT      13-APR-2000.
FT      CDS
FT      09-OCT-1998; 98DE-1046493.
FT      09-OCT-1998; 98DE-1046493.
FT      PA
FT      (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
FT      Beyer S, Mueller R;
FT      DR
FT      WPI; 2000-294101/26.
FT      CDS
FT      DNA sequence coding for products involved in the biosynthesis of
FT      polypeptide or heteropolypeptide compounds, especially epothilone
FT      PS
FT      Claim 3; Page 20-33; 36pp; German.
FT      XX
FT      This invention describes a novel DNA sequence (1) whose expression
FT      CC products effect or are involved in the enzymatic biosynthesis,
FT      CC mutasynthesis or partial synthesis of polypeptide or heteropolypeptide
FT      CC compounds (II). (1) can be inserted into an expression vector and used
FT      CC to transform or transfect prokaryotic or eukaryotic cells with the aim
FT      CC of obtaining strains that produce large amounts of polypeptide or
FT      CC heteropolypeptide compounds, especially epothilones, which have cytotoxic
FT      CC and/or immunosuppressant and antibiotic and antifungal activities and
FT      CC are useful as plant-protection agents. This sequence represents the DNA
FT      CC sequence isolated from Sorangium cellulosum which is described in the
FT      CC method of the invention.
FT      XX
FT      Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 other;

```

```

Oy      3 gtcaggtcaggtc 16
Db      35785 gtcaggtcaggtc 35798

Query Match      56.08; Score 14; DB 21; Length 37856;
Best Local Similarity 100.08; Pred. No. 1e-02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 40
AAV30458
ID      AAV30458 standard; DNA: 534720 BP.
AC      AAV30458;
DT      14-OCT-1998 (first entry)
XX      Rhizobium species plasmid pNGR234a.
DE      Rhizobium species plasmid pNGR234a.
XX      Symbiosis: open reading frame; ORF; plasmid; vector; transportation;
KW      degradation; metabolism; host range; nitrogen fixation; nodulation;
KW      legume; plant; ds.
XX      Rhizobium sp.
OS      Rhizobium sp.
FH      Key
FT      CDS
FT      /location/qualifiers
FT      417796..418671
FT      /tag= a
FT      /standard_name= "ORF K1"
FT      /product= "oligopeptide permease"
FT      /note= "homologous to the Oppc gene"
FT      418673..419680
FT      /tag= b
FT      /standard_name= "ORF K2"
FT      /product= "oligopeptide permease"
FT      /note= "homologous to the Oppd gene"
FT      419677..420738
FT      /tag= c
FT      /standard_name= "ORF K3"
FT      /product= "oligopeptide permease"
FT      /note= "homologous to the Oppf gene"
FT      420774..422159
FT      /tag= d
FT      /standard_name= "ORF K4"
FT      /product= "encapsulation-like protein"
FT      /note= "homologous to the Capa gene"
FT      422628..424031
FT      /tag= e
FT      /standard_name= "ORF K5"
FT      /product= "aminotransferase-like protein"
FT      /note= "homologous to the BioA gene"
FT      424056..425594
FT      /tag= f
FT      /standard_name= "ORF K6"
FT      /product= "(semi)aldehyde dehydrogenase-like protein"
FT      complement (426949..428028)
FT      /tag= g
FT      /standard_name= "ORF K7"
FT      /product= "transposase homologue"
FT      /note= "homologous to the Trp gene"
FT      428292..429623
FT      /tag= h
FT      /standard_name= "ORF K8"
FT      /product= "glutamate dehydrogenase-like protein"
FT      /note= "homologous to the GUD1 gene"
FT      complement (430538..431284)
FT      /tag= i
FT      /standard_name= "ORF K9"
FT      /product= "13:isopase homologue"
FT      complement (431296..432840)
FT      /tag= j
FT      CDS

```

```

FT      /standard_name= "ORF K10"
FT      /product= "transposase homologue"
FT      /note= "homologous to the Tnp gene"
FT      complement (433880..434110)
FT      /tag= k
FT      /standard_name= "ORF K11"
FT      /product= "protein of unknown function"
FT      /note= "homologous to the fixu gene"
FT      complement (434107..434433)
FT      /tag= l
FT      /standard_name= "ORF K12"
FT      /product= "protein of unknown function"
FT      complement (434517..434711)
FT      /tag= m
FT      /standard_name= "ORF K13"
FT      /product= "ferrodoxin/ferrodoxin-like protein"
FT      /note= "homologous to the fdxn gene"
FT      complement (434753..436234)
FT      /tag= n
FT      /standard_name= "ORF K14"
FT      /gene= "nifB"
FT      /product= "protein involved in Feko co-factor
FT      biosynthesis"
FT      complement (436460..438130)
FT      /tag= o
FT      /standard_name= "ORF K15"
FT      /gene= "nifA"
FT      /product= "positive regulator of nif, fix and other
FT      genes"
FT      complement (438297..438590)
FT      /tag= p
FT      /standard_name= "ORF K16"
FT      /gene= "fixX"
FT      /product= "protein required for nitrogenase activity"
FT      complement (438605..439912)
FT      /tag= q
FT      /standard_name= "ORF K17"
FT      /gene= "fixC"
FT      /product= "protein required for nitrogenase activity"
FT      complement (439923..441032)
FT      /tag= r
FT      /standard_name= "ORF K18"
FT      /gene= "fixB"
FT      /product= "protein required for nitrogenase activity"
FT      complement (441042..441899)
FT      /tag= s
FT      /standard_name= "ORF K19"
FT      /gene= "fixA"
FT      /product= "protein required for nitrogenase activity"
FT      complement (442316..4442636)
FT      /tag= t
FT      /standard_name= "ORF K20"
FT      /product= "protein of unknown function"
FT      complement (443313..443879)
FT      /tag= u
FT      /standard_name= "ORF K21"
FT      /product= "protein of unknown function"
FT      complement (444337..445029)
FT      /tag= v
FT      /standard_name= "ORF K22"
FT      /product= "ferrodoxin-like protein"
FT      /note= "homologous to the NifH gene"
FT      complement (445088..446602)
FT      /tag= w
FT      /standard_name= "ORF K23"
FT      /gene= "dctA"
FT      /product= "C4-dicarboxylate transport protein"
FT      /note= "homologous to the DctAI gene"
FT      complement (446599..447843)
FT      /tag= x
FT      /standard_name= "ORF l1"
FT      /product= "cytochrome P450-like protein"
FT      /note= "homologous to the CamC gene"
FT      complement (447844..448500)
FT      /tag= y
FT      /standard_name= "ORF L2"
FT      /product= "gamma-hexachlorocyclohexane-dechlorinase-like
FT      protein"
FT      complement (448497..450203)
FT      /tag= z
FT      /standard_name= "ORF L3"
FT      /product= "putative protein with degradative function"
FT      complement (450341..451396)
FT      /tag= aa
FT      /standard_name= "ORF L4"
FT      /product= "luciferase alpha-subunit-like protein"
FT      /note= "homologous to the luxA gene"
FT      complement (452980..454494)
FT      /tag= ab
FT      /standard_name= "ORF L6"
FT      /gene= "nifD"
FT      /product= "alpha-subunit of Feko protein of nitrogenase"
FT      complement (454590..456131)
FT      /tag= ac
FT      /standard_name= "ORF L7"
FT      /gene= "nifK"
FT      /product= "beta-subunit of Feko protein of nitrogenase"
FT      complement (456187..457677)
FT      /tag= ad
FT      /standard_name= "ORF L8"
FT      /product= "protein involved in Feko co-factor
FT      biosynthesis"
FT      complement (457687..459096)
FT      /tag= ae
FT      /standard_name= "ORF L9"
FT      /product= "protein involved in Feko co-factor
FT      biosynthesis"
FT      complement (459093..459575)
FT      /tag= af
FT      /standard_name= "ORF L10"
FT      /product= "protein of unknown function"
FT      /note= "homologous to the Nifx gene"
FT      complement (459579..460067)
FT      /tag= ag
FT      /standard_name= "ORF L11"
FT      /product= "protein of unknown function"
FT      complement (460501..460920)
FT      /tag= ah
FT      /standard_name= "ORF L12"
FT      /product= "protein similar to part of the Fe protein
FT      of nitrogenase"
FT      complement (461228..461545)
FT      /tag= ai
FT      /standard_name= "ORF L13"
FT      /product= "protein of unknown function"
FT      complement (463201..464739)
FT      /tag= aj
FT      /standard_name= "ORF L14"
FT      /product= "peptidase-like protein"
FT      /note= "homologous to the Df-MPP gene"
FT      complement (464736..466079)
FT      /tag= ak
FT      /standard_name= "ORF L15"
FT      /product= "processing protease-like protein"
FT      /note= "homologous to the pp gene"
FT      complement (466590..467021)

```

```

Query Match      56.08; Score 14; DB 19; Length 534720;
Best Local Similarity 100.08; Pred. NO. 89;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      4 tcaagtcaggtca 17

```

Db 435493 tcaggtcaggtca 435506
 |||

RESULT 41

AAV30459 standard; DNA: 536165 BP.

AAV30459;

06-JUL-1999 (first entry)

Rhizobium species symbiotic plasmid pNCR234.

Symbiosis: open reading frame; ORF; plasmid; vector; transportation;
 degradation; metabolism; host range; nitrogen fixation; nodulation;
 legume; plant; ds.

Rhizobium sp.

MO9802560-A2.

22-JAN-1998.

10-JUL-1997; 97WO-1800950.

20-MAY-1997; 97GB-0010395.

12-JUL-1996; 96EP-0730001.

(MOLE-) INST MOLECULAR BIOTECHNOLOGY

(BIOL-) LAB BIOLOGIE MOLEculaire PLANTES SUPERIE.

Broughton WJ, Freiberg CB, Perret XP, Rosenthal A;

WPI: 1998-110606/10.

New isolated symbiotic plasmid from Rhizobium sp. NCR234 - used to
 develop products for modifying plant characteristics, e.g. nitrogen
 fixation, synthesis of compounds and stress response

Claim 1; Fig 3; 228pp; English.

This is the nucleotide sequence of the plasmid pNCR234 isolated from
 Rhizobium sp. NCR234. Open reading frames (ORF) derivable from the
 nucleotide sequence are claimed. The nucleotide sequences or ORFs can
 be used e.g. in the transportation of compounds to and from an organism
 which is a host to at least one of the nucleotide sequences, ORFs or
 proteins, the degradation and/or metabolism of organic, inorganic,
 natural or xenobiotic substances in a host organism or the modification
 of the host range, nitrogen fixation abilities for obtaining a synthetic
 minimal set of ORFs required for functional Rhizobium-legume symbiosis,
 especially for nodulation efficiency on host plants.

Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 536165;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcaggtcaggtca 17

Db 435493 tcaggtcaggtca 435506

RESULT 42

AAO13647 standard; DNA: 18 BP.

AAO13647;

28-NOV-1991 (first entry)

Exon B/intron junction of PTPase B1 gene.

PTPBI; protein phosphotyrosyl phosphatase B1;
 growth suppression activity; ANLL; myeloid proliferation; ss.

Homo sapiens.

Key Location/Qualifiers

exon

intron

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

1.9

/tag= a

/note= "last 3 codons of exon B"

10..18

/tag= b

/note= "5' end of intron"

WO9113173-A.

05-SEP-1991.

01-MAR-1991; 91WO-0001432.

18-JAN-1991; 91US-0643041.

02-MAR-1990; 90US-0487733.

(BIOT-) APPL BIOTECHN INC.

Bruskin AM, Hill DE;

WPI: 1991-281485/38.

Vector contg. protein phospho:tyrosyl phosphatase gene fragment -

for use in diagnosis of cancers, e.g. acute non-lymphocytic

leukemia

Disclosure: Fig 4B; 73pp; English.

The PTPase B1 genomic sequence comprises a number of exons and

introns. Exon B encodes amino acids Met(235) to Gln(288) of the

PTPase B1 protein (see ARI3907). This is the junction of the 3' end

of exon B and the 5' end of the intron.

See AAO13643-Q13654.

Sequence 18 BP; 3 A; 5 C; 5 G; 5 T; 0 other;

Query Match 52.0%; Score 13; DB 12; Length 18;

Best Local Similarity 100.0%; Pred. No. 4.7e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gtgcagtcagca 20

Db 4 gtgcagtcagca 16

RESULT 43

AA85925/c

AA85925 standard; DNA: 123 BP.

AA85925;

13-SEP-1999 (first entry)

Oligonucleotide used to produce heavy chain variable region of Ig NEM.

Light chain variable region; interleukin-4; IL-4; antibody 3B9;

chimeric antibody; humanised antibody; IL-4 mediated allergic reaction;

Immunoglobulin E-mediated allergic reaction; allergic rhinitis;

conjunctivitis; atopic dermatitis; atopic asthma; anaphylactic shock;

rheumatoid arthritis; host-versus-graft disease; renal disease;

allergy; Ig NEW; ss.

Synthetic.

```

PN  US5928904-A.
XX
PD  27-JUL-1999.
XX
XX
PF  07-JUN-1995; 95US-0483632.
XX
PR  07-JUN-1995; 95US-0483632.
PR  07-SEP-1993; 93US-0117356.
PR  14-OCT-1993; 93US-0136783.
PR  07-SEP-1994; 94WO-US10308.
XX
PA  (SMIK ) SMITHKLINE BEECHAM CORP.
XX  (SMIK ) SMITHKLINE BEECHAM PLC.
PI  Gross MS, Holmes SD, Sylvester DR;
XX
DR  WPI: 1999-429500/36.
XX
PT  New DNA molecules encoding recombinant antibodies useful for
PT  treating IL4-mediated conditions
XX
PS  Disclosure; Columns 61-63; 50pp; English.
XX
CC  The specification describes chimeric and humanised IL-4
CC  monoclonal antibodies. The antibodies of the invention are used in
CC  therapeutic and pharmaceutical compositions for treating IL-4 mediated
CC  and immunoglobulin E-mediated allergic reactions e.g. allergic rhinitis,
CC  conjunctivitis, atopic dermatitis, atopic asthma, anaphylactic shock,
CC  rheumatoid arthritis, host-versus-graft disease and renal disease.
CC  They are also useful in the diagnosis of an allergy or condition
CC  associated with excess IL-4 production through the measurement e.g. by
CC  ELISA of circulating endogenous IL-4 levels in humans. Oligonucleotides
CC  AAX85925-28 were used to produce the heavy chain variable region of
CC  Ig NEM, in the course of the invention.
XX
SQ  Sequence 123 BP; 25 A; 37 C; 41 G; 20 T; 0 other;

Query Match          52.0%; Score 13; DB 20; Length 123;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  6 aggtgcaggtcag 18
    |||
DB  70 AGGTGCAGGTCAG 58

RESULT 44
AII23772
ID  AII23772 standard; DNA; 153 BP.
XX
AC  AII23772;
XX
DT  12-OCT-2001 (first entry)
XX
DE  Probe #13705 for gene expression analysis in human cervical cell sample.
XX
KW  Probe; human; microarray; gene expression; cervical epithelial cell;
KW  cervical cancer; ss.
XX
OS  Homo sapiens.
XX
PN  WO200157278-A2.
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00670.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  04-OCT-2000; 2000US-0234687.

```

```

PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI: 2001-488901/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human cervical epithelial cells -
XX
PS  Claim 25; SEQ ID No 13705; 487pp; English.
XX
CC  The present invention relates to human single exon nucleic acid probes
CC  (SENP). The present sequence is one such probe. The SENPs are derived
CC  from human HeLa cells. The SENPs can be used to produce a single exon
CC  microarray, which can be used for measuring human gene expression in a
CC  sample derived from human cervical epithelial cells. By measuring gene
CC  expression, the probes are therefore useful in grading and/or staging
CC  of diseases of the cervix, notably cervical cancer.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 153 BP; 33 A; 61 C; 29 G; 29 T; 1 other;

Query Match          52.0%; Score 13; DB 22; Length 153;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  9 tgcagtcagcac 21
    |||
DB  4 tgcagtcagcac 16

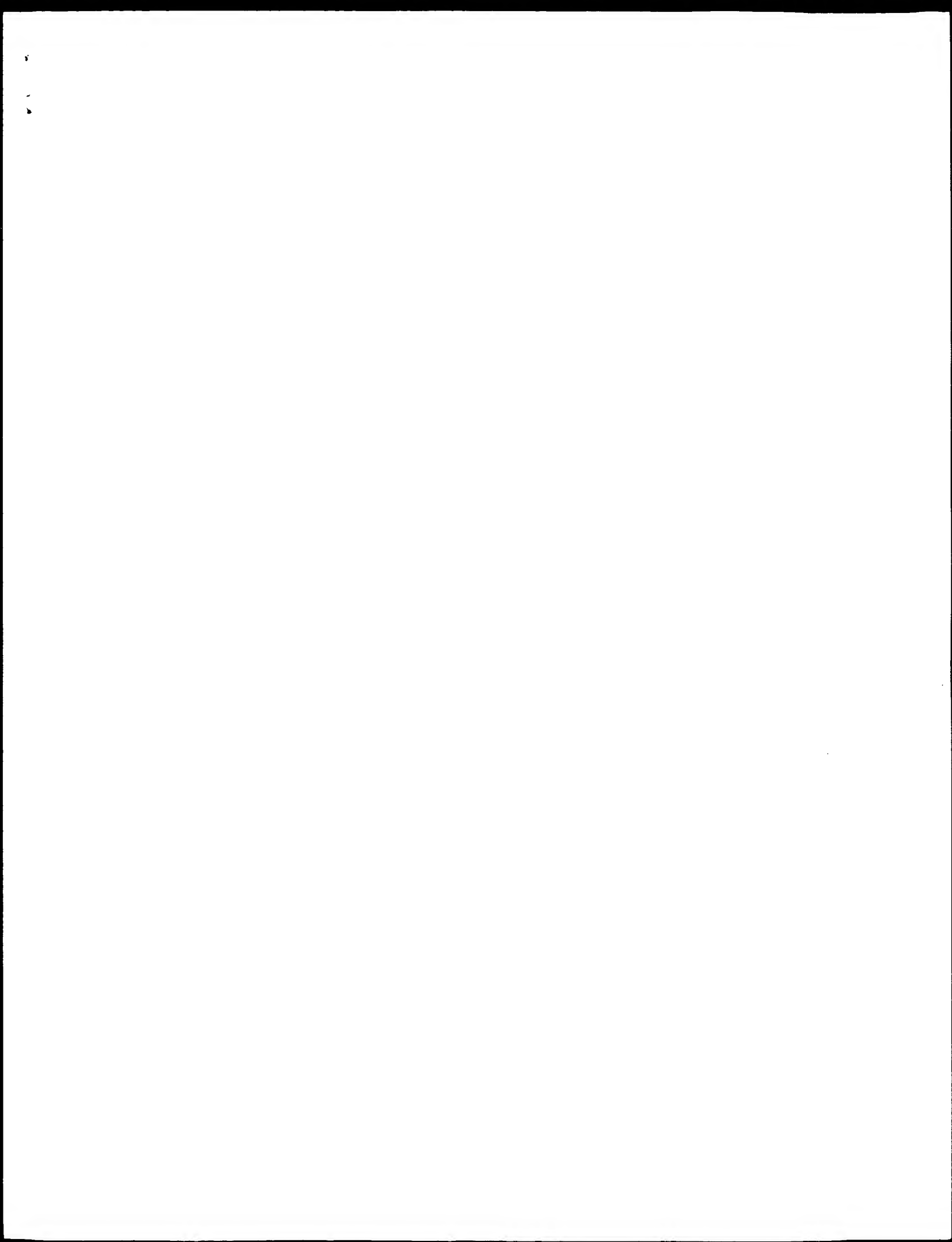
RESULT 45
AII49080
ID  AII49080 standard; DNA; 153 BP.
XX
AC  AII49080;
XX
DT  17-OCT-2001 (first entry)
XX
DE  Probe #17766 used to measure gene expression in human placenta sample.
XX
KW  Probe; microarray; human; placenta; antenatal diagnosis;
KW  genetic disorder; ss.
XX
OS  Homo sapiens.
XX
PN  WO200157272-A2.
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00663.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  04-OCT-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI: 2001-48897/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for

```

PR analyzing gene expression in human placenta -
 XX
 PS Claim 25: SEQ ID No 17766; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 153 BP; 33 A; 61 C; 29 G; 29 T; 1 other;

Query Match 52.0%; Score 13; DB 22; Length 153;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 tgcaggtcagcac 21
 |||||
 Db 4 tgcaggtcagcac 16

Search completed: December 26, 2001, 14:17:19
 Job time: 7031 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:55:41 : Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196f-10

Perfect score: 25

Sequence: 1 tgcgcagtcagtcagtcagcttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 13726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

1: /cgn2_6/prodata/2/ina/5a_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5b_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5a_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5b_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2 US-08-401-068-7	Sequence 7, Appl1
2	25	100.0	1041	2 US-08-846-938-7	Sequence 7, Appl1
3	25	100.0	5872	3 US-08-411-768B-1	Sequence 1, Appl1
4	25	100.0	5872	3 US-08-411-768B-6	Sequence 6, Appl1
5	15	60.0	630	3 US-08-545-809A-10	Sequence 10, Appl1
6	14	56.0	613	3 US-08-545-809A-5	Sequence 5, Appl1
7	14	56.0	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
8	14	56.0	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
9	14	56.0	4411529	4 US-09-103-840A-1	Sequence 1, Appl1
10	13	52.0	123	2 US-08-483-636-59	Sequence 59, Appl1
11	13	52.0	123	2 US-08-483-632-59	Sequence 59, Appl1
12	13	52.0	363	2 US-08-483-632-63	Sequence 63, Appl1
13	13	52.0	363	2 US-08-483-632-63	Sequence 63, Appl1
14	13	52.0	378	1 US-08-488-376-13	Sequence 13, Appl1
15	13	52.0	378	1 US-08-488-376-13	Sequence 13, Appl1
16	13	52.0	378	2 US-08-634-223-13	Sequence 13, Appl1
17	13	52.0	378	2 US-08-634-223-13	Sequence 13, Appl1
18	13	52.0	378	2 US-08-634-223-13	Sequence 13, Appl1
19	13	52.0	378	2 US-08-634-224-13	Sequence 13, Appl1
20	13	52.0	378	2 US-08-634-224-13	Sequence 13, Appl1
21	13	52.0	378	2 US-08-634-400-13	Sequence 13, Appl1
22	13	52.0	378	2 US-08-634-400-13	Sequence 13, Appl1
23	13	52.0	378	2 US-08-635-878-13	Sequence 13, Appl1
24	13	52.0	378	2 US-08-635-878-13	Sequence 13, Appl1
25	13	52.0	378	2 US-08-770-057-13	Sequence 13, Appl1
26	13	52.0	378	2 US-08-770-057-13	Sequence 13, Appl1
27	13	52.0	378	4 US-09-335-697B-13	Sequence 13, Appl1

c 28	13	52.0	423	2	US-08-345-321-9	Sequence 9, Appl1
c 29	13	52.0	516	3	US-08-545-809A-26	Sequence 26, Appl1
c 30	13	52.0	748	3	US-09-154-083-31	Sequence 31, Appl1
c 31	13	52.0	750	4	US-09-183-959-18	Sequence 18, Appl1
c 32	13	52.0	879	3	US-08-714-071-3	Sequence 3, Appl1
c 33	13	52.0	1040	4	US-09-183-959-11	Sequence 11, Appl1
c 34	13	52.0	1173	3	US-08-706-216-5	Sequence 5, Appl1
c 35	13	52.0	1428	1	US-08-488-376-17	Sequence 17, Appl1
c 36	13	52.0	1428	1	US-08-488-376-19	Sequence 19, Appl1
c 37	13	52.0	1428	2	US-08-634-223-17	Sequence 17, Appl1
c 38	13	52.0	1428	2	US-08-634-223-19	Sequence 19, Appl1
c 39	13	52.0	1428	2	US-08-634-224-17	Sequence 17, Appl1
c 40	13	52.0	1428	2	US-08-634-224-19	Sequence 19, Appl1
c 41	13	52.0	1428	2	US-08-634-400-17	Sequence 17, Appl1
c 42	13	52.0	1428	2	US-08-634-400-19	Sequence 19, Appl1
c 43	13	52.0	1428	2	US-08-635-878-17	Sequence 17, Appl1
c 44	13	52.0	1428	2	US-08-635-878-19	Sequence 19, Appl1
c 45	13	52.0	1428	2	US-08-770-057-17	Sequence 17, Appl1
c 46	13	52.0	1428	2	US-08-770-057-19	Sequence 19, Appl1
c 47	13	52.0	1428	4	US-09-335-697B-17	Sequence 17, Appl1
c 48	13	52.0	1428	4	US-09-335-697B-19	Sequence 19, Appl1
c 49	13	52.0	1640	4	US-08-676-444-41	Sequence 41, Appl1
c 50	13	52.0	2830	2	US-09-010-928B-1	Sequence 1, Appl1
c 51	13	52.0	3105	4	US-08-542-635-1	Sequence 1, Appl1
c 52	13	52.0	4465	2	US-08-620-605D-1	Sequence 1, Appl1
c 53	13	52.0	4533	3	US-08-726-214-5	Sequence 5, Appl1
c 54	13	52.0	4547	2	US-09-005-232A-1	Sequence 1, Appl1
c 55	13	52.0	15239	1	US-08-390-878-17	Sequence 17, Appl1
c 56	13	52.0	49272	1	US-08-614-770A-1	Sequence 1, Appl1
c 57	13	48.0	19	1	US-07-936-421-11	Sequence 11, Appl1
c 58	12	48.0	20	3	US-08-545-809A-72	Sequence 72, Appl1
c 59	12	48.0	28	2	US-08-353-372A-2	Sequence 2, Appl1
c 60	12	48.0	40	2	US-08-805-918-11	Sequence 11, Appl1
c 61	12	48.0	40	2	US-08-811-028-7	Sequence 7, Appl1
c 62	12	48.0	59	2	US-08-116-778E-24	Sequence 24, Appl1
c 63	12	48.0	59	2	US-08-438-562-24	Sequence 24, Appl1
c 64	12	48.0	59	2	US-08-483-528B-24	Sequence 24, Appl1
c 65	12	48.0	59	3	US-08-673-799C-24	Sequence 24, Appl1
c 66	12	48.0	64	1	US-08-388-672A-8	Sequence 8, Appl1
c 67	12	48.0	64	3	US-09-080-554-8	Sequence 8, Appl1
c 68	12	48.0	66	1	US-08-207-996-5	Sequence 5, Appl1
c 69	12	48.0	66	2	US-08-760-840A-5	Sequence 5, Appl1
c 70	12	48.0	66	3	US-09-266-119-5	Sequence 5, Appl1
c 71	12	48.0	70	1	US-08-290-592E-39	Sequence 39, Appl1
c 72	12	48.0	70	5	PCTN-US96-09448-39	Sequence 39, Appl1
c 73	12	48.0	75	1	US-08-657-012-5	Sequence 5, Appl1
c 74	12	48.0	75	3	US-09-013-872-5	Sequence 5, Appl1
c 75	12	48.0	75	4	US-09-184-198-5	Sequence 5, Appl1
c 76	12	48.0	76	4	US-09-025-769B-134	Sequence 134, Appl1
c 77	12	48.0	93	4	US-08-556-978B-67	Sequence 67, Appl1
c 78	12	48.0	93	4	US-08-556-978B-68	Sequence 68, Appl1
c 79	12	48.0	138	1	US-08-467-420A-58	Sequence 58, Appl1
c 80	12	48.0	138	1	US-08-470-110A-58	Sequence 58, Appl1
c 81	12	48.0	138	1	US-08-667-769A-58	Sequence 58, Appl1
c 82	12	48.0	138	5	US-08-940-371-58	Sequence 58, Appl1
c 83	12	48.0	138	5	PCTN-US95-17082A-58	Sequence 58, Appl1
c 84	12	48.0	152	1	US-08-136-277-9	Sequence 9, Appl1
c 85	12	48.0	152	2	US-08-479-403-9	Sequence 9, Appl1
c 86	12	48.0	152	3	US-08-835-734-9	Sequence 9, Appl1
c 87	12	48.0	162	1	US-08-275-916-1	Sequence 1, Appl1
c 88	12	48.0	162	5	PCTN-US93-03868-1	Sequence 1, Appl1
c 89	12	48.0	198	2	US-08-414-657D-15	Sequence 15, Appl1
c 90	12	48.0	198	2	US-08-414-657D-15	Sequence 15, Appl1
c 91	12	48.0	201	2	US-08-222-719-8	Sequence 8, Appl1
c 92	12	48.0	201	2	US-08-470-925-8	Sequence 8, Appl1
c 93	12	48.0	201	2	US-08-471-613-8	Sequence 8, Appl1
c 94	12	48.0	201	5	PCTN-US93-10443-8	Sequence 8, Appl1
c 95	12	48.0	303	4	US-08-556-978B-82	Sequence 82, Appl1
c 96	12	48.0	357	1	US-08-459-310-11	Sequence 11, Appl1
c 97	12	48.0	357	1	US-08-467-420A-61	Sequence 61, Appl1
c 98	12	48.0	357	1	US-08-470-110A-61	Sequence 61, Appl1
c 99	12	48.0	357	1	US-08-667-769A-61	Sequence 61, Appl1
c 100	12	48.0	357	2	US-08-940-371-61	Sequence 61, Appl1

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 6.6e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcagtgacagtcacagctg 25
|||||
Db 111 TCGTCAGTGCAGTGCAGCAGCTTG 135

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 6.6e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcagtgacagtcacagctg 25
|||||
Db 111 TCGTCAGTGCAGTGCAGCAGCTTG 135

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
```



```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start= 2295
OTHER INFORMATION: /function= "Involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /EC_number= 2.6.1.62
OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
OTHER INFORMATION: aminotransf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "DioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match          100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcagtcgacgacgacgtg 25
   |||||
Db 227 tcgtcagtcgacgacgacgtg 251

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: PRO30A15-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1154..2308
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start=1154
; OTHER INFORMATION: /EC_number=2.3.1.47
; OTHER INFORMATION: /product="KAPA synthase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /gene="bioF"
; OTHER INFORMATION: /number=2
; OTHER INFORMATION: /standard_name="8-Amino-7-oxononanoate synthase"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3043..3753
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start=3043
; OTHER INFORMATION: /EC_number=6.3.3.3
; OTHER INFORMATION: /product="DTPB synthase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /gene="bioid"
; OTHER INFORMATION: /number=4
; OTHER INFORMATION: /standard_name="Dethiobiotin synthase"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 1141..1156
; OTHER INFORMATION: /standard_name="bioF RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3030..3045
; OTHER INFORMATION: /standard_name="bioid RBS"
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
; US-08-411-768B-6

Query Match          100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcgtcagtgacggtcacgcttg 25
    |||||||||||||||||||
Db 227 tcgtcagtgacggtcacgcttg 251

RESULT 5
US-08-545-809A-10/c
; Sequence 10, Application US/08545809A
; Patent No. 6096878
```

```

; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Matsuda, Fumihiko
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 145
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,809A
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/00603
; FILING DATE: 10-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 06501/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: human lymphoblast
; CELL LINE: CGM1
; US-08-545-809A-10

Query Match          60.0%; Score 15; DB 3; Length 630;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 aggtgcaggtcacga 20
    |||||||||||||||
Db 280 aggtgcaggtcacga 266

RESULT 6
US-08-545-809A-5/c
; Sequence 5, Application US/08545809A
; Patent No. 6096878
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Matsuda, Fumihiko
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 145
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,809A
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SFO ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: CGM1
US-08-545-809A-5

Query Match 56.0%; Score 14; DB 3; Length 613;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 aggtcaggtcagc 19
|||||
DB 283 AGTGCAGGTGACG 270

RESULT 7
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match 56.0%; Score 14; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 gtgcaggtcagc 21

|||||
Db 586964 gtgcaggtcagc 586977

RESULT 8
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match 56.0%; Score 14; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 ggtgcaggtcagca 20
|||||
Db 2740489 GGTGCAGGTGACGA 2740476

RESULT 9
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 56.0%; Score 14; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 ggtgcaggtcagca 20
|||||
Db 2743678 GGTGCAGGTGACGA 2743665

RESULT 10
US-08-483-636-59/c
Sequence 59, Application US/08483636
Patent No. 5914110
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
Treatment of IL4 Mediated Disorders
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corp./Corporate
ADDRESS: Intellectual Property
STREET: P.O. Box 1539 / UW2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,636
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-483-636-59

Query Match 52.0%; Score 13; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcagctcag 18
|||||
DB 70 AGGTGCAGCTCAG 58

RESULT 11
US-08-483-632-59/c
Sequence 59, Application US/08483632
Patent No. 5928904
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
Treatment of IL4 Mediated Disorders

NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corp./Corporate
ADDRESS: Intellectual Property
STREET: P.O. Box 1539 / UW2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,632
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-483-632-59

Query Match 52.0%; Score 13; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcagctcag 18
|||||
DB 70 AGGTGCAGCTCAG 58

RESULT 12
US-08-483-636-63/c
Sequence 63, Application US/08483636
Patent No. 5914110
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
Treatment of IL4 Mediated Disorders
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corp./Corporate
ADDRESS: Intellectual Property
STREET: P.O. Box 1539 / UW2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,636
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-636-63

Query Match 52.0%; Score 13; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 aggtgcaggtcag 18
|||||
Db 70 AGGTGCAGGTGAG 58

RESULT 13
US-08-483-632-63/C
Sequence 63 Application US/08483632
Patent No. 5928904
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies useful in
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corp./Corporate
ADDRESS: Intellectual Property
STREET: P.O. Box 1539 / UM2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,632
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366

FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-632-63

Query Match 52.0%; Score 13; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 aggtgcaggtcag 18
|||||
Db 70 AGGTGCAGGTGAG 58

RESULT 14
US-08-488-376-13/C
Sequence 13 Application US/08488376
Patent No. 5811524
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:

LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-488-376-13

Query Match 52.0%; Score 13; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcaggtcagc 19
|||||
Db 69 ggtgcaggtcagc 57

RESULT 15
US-08-488-376-15/c
Sequence 15, Application US/08488376
Patent No. 5811524
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulatma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-488-376-15

Query Match 52.0%; Score 13; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 6 aggtcaggtcagc 18
|||||
Db 70 aggtcaggtcagc 58

RESULT 16
US-08-634-223-13/c
Sequence 13, Application US/08634223
Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulatma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-223-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcaggtcagc 19
|||||
Db 69 ggtgcaggtcagc 57

RESULT 17
US-08-634-223-15/c
Sequence 15, Application US/08634223

Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-223-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcagtcacg 18
|||||
Db 70 AGGTGCAGTCACG 58

RESULT 18
US-08-634-224-13/c
Sequence 13, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-224-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcagtcacg 19
|||||
Db 69 GGTGCAGTCACG 57

RESULT 19
US-08-634-224-15/c
Sequence 15, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-224-15

Query Match      52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 aggtgcaggtcagc 18
      |||
Db      70 AGGTGACGCTCAG 58

RESULT 20
US-08-634-400-13/c
Sequence 13, Application US/08634400
Patent No. 5939068
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,400
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376

```

```

FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-400-13

Query Match      52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 ggtgcaggtcagc 19
      |||
Db      69 GGTGACGCTCAGC 57

RESULT 21
US-08-634-400-15/c
Sequence 15, Application US/08634400
Patent No. 5939068
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,400
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:

```


LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-400-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcag 18
|||||
DB 70 AGGTGCAGGTGAG 58

RESULT 22

US-08-635-878-13/c
Sequence 13, Application US/08635878
Patent No. 5955364

GENERAL INFORMATION:

APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,878
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-635-878-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcaggtcagc 19
|||||
DB 69 GGTGCAGGTGAGC 57

RESULT 23

US-08-635-878-15/c
Sequence 15, Application US/08635878
Patent No. 5955364

GENERAL INFORMATION:

APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,878
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-635-878-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcag 18
|||||
DB 70 AGGTGCAGGTGAG 58

RESULT 24
US-08-770-057-13/C
Sequence 13, Application US/08770057
Patent No. 5958765
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-770-057-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagc 19
|||||
Db 69 ggtgcaggtcagc 57

RESULT 25
US-08-770-057-15/C
Sequence 15, Application US/08770057
Patent No. 5958765
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne

APPLICANT: NEMMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-770-057-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtcaggtcagc 18
|||||
Db 70 aggtcaggtcagc 58

RESULT 26
US-09-335-697B-13/C
Sequence 13, Application US/09335697B
Patent No. 6200804
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEMMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia

COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,697B
FILING DATE: 06-Jul-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/770,057
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-335-697B-13

Query Match 52.0%; Score 13; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagc 19
|||||
DB 69 ggtgcaggtcagc 57

RESULT 27
US-09-335-697B-15/c
Sequence 15, Application US/09335697B
Patent No. 6200804
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
CHAMAT, Soulaime Salim
PAN, Li-zhen
WALSH, Edward E.
HEARD, Cheryl Janne
NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,697B

FILING DATE: 06-Jul-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/770,057
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-335-697B-15

Query Match 52.0%; Score 13; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcagc 18
|||||
DB 70 aggtgcaggtcagc 58

RESULT 28
US-08-345-321-9/c
Sequence 9, Application US/08345321
Patent No. 5914109
GENERAL INFORMATION:
APPLICANT: ZOLLA-PAZNER, Susan
TITLE OF INVENTION: HETEROHYBRIDOMAS PRODUCING HUMAN
MONOCLONAL ANTIBODIES TO HIV-1
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/345,321
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/872,675
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: ZOLLA-PAZNER1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..423
US-08-345-321-9

Query Match 52.0%; Score 13; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 aggtcagctcag 18
|||||
Db 127 AGGTCAAGTCAAG 115

RESULT 29

US-08-545-809A-26/C
Sequence 26, Application US/08545809A
Patent No. 6096878

GENERAL INFORMATION:

APPLICANT: Honjo, Tasuku

APPLICANT: Matsuda, Fumihiko

TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE

TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME

NUMBER OF SEQUENCES: 145

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FASTED for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/545,809A

FILING DATE: 27-MAR-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/00603

FILING DATE: 10-MAY-1993

ATTORNEY/AGENT INFORMATION:

NAME: Freeman, John W.

REGISTRATION NUMBER: 29,066

REFERENCE/DOCKET NUMBER: 06501/004001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-542-5070

TELEFAX: 617-542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 546 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

CELL TYPE: human lymphoblast

CELL LINE: CGM1

US-08-545-809A-26

Query Match 52.0%; Score 13; DB 3; Length 546;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 7 ggtcagctcagc 19
|||||
Db 232 GGTCAGCTCAGC 220

RESULT 30

US-09-154-083-31

Sequence 31, Application US/09154083

Patent No. 6150513

GENERAL INFORMATION:

APPLICANT: Wu, Kai

TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA

TITLE OF INVENTION: Constructs therefor

FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz

CURRENT APPLICATION NUMBER: US/09/154,083

CURRENT FILING DATE: 1998-09-16

NUMBER OF SEQ ID NOS: 31

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 31

LENGTH: 718

TYPE: DNA

ORGANISM: Streptomyces hygroscopicus

US-09-154-083-31

Query Match 52.0%; Score 13; DB 3; Length 718;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 gcagtcagcagc 22
|||||
Db 347 gcagtcagcagc 359

RESULT 31

US-09-183-959-18

Sequence 18, Application US/09183959

Patent No. 6303332

GENERAL INFORMATION:

APPLICANT: Cahoon, Rebecca E.

APPLICANT: Miao, Guo-Hua

APPLICANT: Rafalski, J. Antoni

APPLICANT: Tarantino, Graziana

TITLE OF INVENTION: CORN CDNA ENCODING SOUTHERN LEAF BLIGHT RESISTANCE

FILE REFERENCE: BB-1125

CURRENT APPLICATION NUMBER: US/09/183,959

CURRENT FILING DATE: 1998-11-02

EARLIER APPLICATION NUMBER: 60/064,493

EARLIER FILING DATE: No. 6303332ember 5, 1997

NUMBER OF SEQ ID NOS: 20

SOFTWARE: Microsoft Windows 95

SEQ ID NO 18

LENGTH: 750

TYPE: DNA

ORGANISM: Zea mays

US-09-183-959-18

Query Match 52.0%; Score 13; DB 4; Length 750;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 ggtcagctcagc 19
|||||
Db 687 ggtcagctcagc 699

RESULT 32

US-08-714-071-3/C

Sequence 3, Application US/08714071

Patent No. 6136584

GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Shiro, OKUNO
APPLICANT: Hisanobu, HIRANO
APPLICANT: Sadahito, SHIN
TITLE OF INVENTION: FK506 BINDING PROTEIN GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,071
CLASSIFICATION: 435
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 879 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Human fetal brain cDNA library
IMMEDIATE SOURCE:
CLONE: OTK4(6-1)
FEATURE:
NAME/KEY: CDS
LOCATION: 70..393
US-08-714-071-3

Query Match 52.0%; Score 13; DB 3; Length 879;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagc 19
|||||

DB 303 GGTGCAGGTGACG 291

RESULT 33
US-09-183-959-11
Sequence 11, Application US/09183959
Patent No. 6303352
GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Miao, Guo-Hua
APPLICANT: Rafalski, J. Antoni
APPLICANT: Taramino, Graziana
TITLE OF INVENTION: CORN CDNA ENCODING SOUTHERN LEAF BLIGHT RESISTANCE
FILE REFERENCE: BB-1125
CURRENT APPLICATION NUMBER: US/09/183,959
CURRENT FILING DATE: 1998-11-02
EARLIER APPLICATION NUMBER: 60/064,493
EARLIER FILING DATE: No. 6303332ember 5, 1997
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Microsoft Windows 95
SEQ ID NO 11

LENGTH: 1040
TYPE: DNA
ORGANISM: Zea mays
US-09-183-959-11

Query Match 52.0%; Score 13; DB 4; Length 1040;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagc 19
|||||

DB 426 ggtgcaggtcagc 438

RESULT 34
US-08-706-216-5/c
Sequence 5, Application US/08706216
Patent No. 6140098
GENERAL INFORMATION:
APPLICANT: Balasubramanian, Sriram
APPLICANT: Ford, John
APPLICANT: Gorman, Daniel M.
APPLICANT: Zurawski, Gerard
TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,216
FILING DATE: 30-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0613
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1173 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1173
US-08-706-216-5

Query Match 52.0%; Score 13; DB 3; Length 1173;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcaggtcaggtc 16
|||||

DB 418 TCAGGTGACGTC 406

RESULT 35
US-08-488-376-17/c

```
Sequence 17, Application US/08488376
Patent No. 5811524
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-488-376-17

Query Match          52.0%; Score 13; DB 1; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 ggtgcagtcagc 19
        |||
Db      126 GGTGCAGTCAGC 114

RESULT 36
US-08-488-376-19/c
Sequence 19, Application US/08488376
Patent No. 5811524
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
```

```
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-488-376-19

Query Match          52.0%; Score 13; DB 1; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 aggtcagtcagc 18
        |||
Db      127 AGGTGCAGTCAGC 115

RESULT 37
US-08-634-223-17/c
Sequence 17, Application US/08634223
Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRICER APPLICATION DATA: US/08/488,376
APPLICATION NUMBER: 07-JUN-1995
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-223-17

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcaggtcag 19
|||||
Db 126 ggtgcaggtcag 114

RESULT 38
US-08-634-223-19/c
Sequence 19, Application US/08634223
Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-223-19

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcag 18
|||||
Db 127 aggtgcaggtcag 115

RESULT 39
US-08-634-224-17/c
Sequence 17, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-224-17

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtcaggtcagc 19
|||||
Db 126 GGTGAGGTGAGC 114

RESULT 40
US-08-634-224-19/c
Sequence 19, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-224-19

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtcaggtcag 18
|||||
Db 127 AGGTGAGGTGAGC 115

RESULT 41
US-08-634-400-17/c
Sequence 17, Application US/08634400
Patent No. 5939068
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,400
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-400-17

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtcaggtcagc 19
|||||
Db 126 GGTGAGGTGAGC 114

RESULT 42
US-08-634-400-19/c
Sequence 19, Application US/08634400
Patent No. 5939068
GENERAL INFORMATION:

APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 2213-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,400
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-400-19

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtcaggtcagc 18
|||||

Db 127 AGGTGACAGTCAG 115

RESULT 43
US-08-635-878-17/c
Sequence 17, Application US/08635878
Patent No. 5955364
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 2213-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,878
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-635-878-17

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ggtcaggtcagc 19
|||||

Db 126 GGTGACAGTCAGC 114

RESULT 44
US-08-635-878-19/c
Sequence 19, Application US/08635878
Patent No. 5955364
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 2213-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/635,878
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/488,376
: FILING DATE: 07-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Teskin, Robin L.
: REGISTRATION NUMBER: 35,030
: REFERENCE/DOCKET NUMBER: 012712-150
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1428 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1428
: US-08-635-878-19

```

```

Query Match      52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      6 aggtcaggtcag 18
      |||
Db      127 AGTGCAGGTGACG 115

```

```

: RESULT 45
: US-08-770-057-17/c
: Sequence 17, Application US/08770057
: Patent No. 5958765
: GENERAL INFORMATION:
: APPLICANT: BRAMS, Peter
: APPLICANT: CHAMAT, Soulaïma Salim
: APPLICANT: PAN, Li-Zhen
: APPLICANT: WALSH, Edward E.
: APPLICANT: HEARD, Cheryl Janne
: APPLICANT: NEWMAN, Roland Anthony
: TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
: TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
: TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/770,057
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/488,376
: FILING DATE: 07-JUN-1995
: ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Teskin, Robin L.
: REGISTRATION NUMBER: 35,030
: REFERENCE/DOCKET NUMBER: 012712-150
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1428 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1428
: US-08-770-057-17

```

```

Query Match      52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      7 ggtgcaggtcagc 19
      |||
Db      126 GGTGCAGGTGACG 114

```

```

Search completed: December 26, 2001, 12:57:44
Job time: 7096 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:10:41 ; Search time 4619.78 Seconds
(Without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196f-10

Perfect score: 25

Sequence: 1 tcgcaggtgcagtcagtcagcttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 9 587657

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: qb_estl:*
11: qb_est2:*
12: qb_hlc:*
13: qb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	17	68.0	324	11	BF464374	UI-M-CGP
2	17	68.0	413	11	BE862745	UI-M-BHO-
3	17	68.0	495	10	A1789420	uk99406.y
4	17	68.0	641	13	A2426129	IM0206H03
5	17	68.0	767	11	BG703880	602686932
6	16	64.0	163	10	AA065941	ml51a01.r
7	16	64.0	288	10	BB516544	BB516544
8	16	64.0	328	11	W54143	W54143
9	16	64.0	338	11	W34110	mc57c06.r1
10	16	64.0	345	10	AV633196	AV633196
11	16	64.0	368	10	AA110457	ml62e03.r
12	16	64.0	370	10	AW822106	uq13c06.x

13	16	64.0	395	11	BF444068	261798.MA
14	16	64.0	398	13	A0298728	HS_3110.B
15	16	64.0	430	10	A1467556	ve3609.x
16	16	64.0	544	10	AW892602	CM3-NN000
17	16	64.0	553	13	A0530462	RPCT-11-3
18	16	64.0	582	13	A2978539	2M0254K21
19	16	64.0	775	10	BE535032	601233227
20	16	64.0	897	13	CNS021CM	AL176719 Tetradon
21	16	64.0	976	11	BG172968	BG172968 602336627
22	16	64.0	1011	12	AK010837	AK010837 Mus muscu
23	16	64.0	1044	13	CNS078T3	AL343413 T7 end of
24	16	64.0	1298	11	BG328084	BG328084 602427134
25	15	60.0	180	10	AA699872	AA699872 zj81f04.s
26	15	60.0	264	11	H22139	H22139 y138a03.s1
27	15	60.0	279	11	R50061	R50061 yj59c10.s1
28	15	60.0	292	10	A1331540	A1331540 fa94b01.y
29	15	60.0	293	13	A2877721	RPCT-23-1
30	15	60.0	300	10	BB175346	BB175346 BB175346
31	15	60.0	306	11	BG100182	BG100182 ux94904.y
32	15	60.0	382	10	A1138808	A1138808 qd98902.x
33	15	60.0	385	10	AW522699	AW522699 UI-R-B00-
34	15	60.0	396	10	AA486040	AA486040 69428.MAR
35	15	60.0	407	10	AA062258	AA062258 ml55908.r
36	15	60.0	407	11	BF599674	BF599674 263546.MA
37	15	60.0	408	13	AA085001	AA085001 HS_2269.A
38	15	60.0	409	11	BF653430	BF653430 277257.MA
39	15	60.0	416	10	AA146146	AA146146 mg65a12.r
40	15	60.0	446	10	AA881231	AA881231 vx10d04.r
41	15	60.0	457	11	BF604067	BF604067 269755.MA
42	15	60.0	466	11	B1398127	B1398127 UNM45G05
43	15	60.0	497	13	AA0921326	AA0921326 RPCT-23-2
44	15	60.0	497	11	BE808281	BE808281 213447.MA
45	15	60.0	495	13	A261270	A261270 RPCT-23-4
46	15	60.0	496	11	BF652819	BF652819 276504.MA
47	15	60.0	525	11	BF567735	BF567735 UI-R-B00-
48	15	60.0	529	10	AA290780	AA290780 MNV046H0
49	15	60.0	548	13	A2887515	A2887515 RPCT-24-1
50	15	60.0	574	11	BF044929	BF044929 BP250009A
51	15	60.0	577	11	BG711675	BG711675 pg11n.pK0
52	15	60.0	580	10	BE262480	BE262480 601145796
53	15	60.0	589	10	AV593752	AV593752 AV593752
54	15	60.0	598	10	AA519851	AA519851 TGE5T269
55	15	60.0	607	13	A2873604	A2873604 IM00187J24
56	15	60.0	620	13	A2342718	A2342718 IM0075120
57	15	60.0	624	13	A2418216	A2418216 IM0194P10
58	15	60.0	639	10	AM342762	AM342762 f186b10.x
59	15	60.0	696	10	AV709400	AV709400 AV709400
60	15	60.0	710	13	A2998889	A2998889 2M0286J03
61	15	60.0	743	10	BE309561	BE309561 601094957
62	15	60.0	804	11	BF864835	BF864835 963055C06
63	15	60.0	811	11	BG965551	BG965551 602830761
64	15	60.0	841	10	BE376456	BE376456 601226248
65	15	60.0	863	11	BF966467	BF966467 602287091
66	15	60.0	946	13	CNS0360N	AL229424 Tetradon
67	15	60.0	984	13	CNS05T79	AL353782 Tetradon
68	15	60.0	990	13	CNS024TL	AL180930 Tetradon
69	15	60.0	1043	13	CNS05049	AL315090 Tetradon
70	15	60.0	1064	11	B1078629	B1078629 602872730
71	15	60.0	1069	13	CNS05GCT	AL336278 Tetradon
72	15	60.0	1101	13	CNS0141H	AL103583 Drosoph11
73	15	60.0	1143	11	BF124974	BF124974 601762117
74	15	60.0	104	11	BF855889	BF855889 PM2-FN021
75	15	60.0	111	11	BF361607	BF361607 POR-L-101
76	15	60.0	143	10	BE066289	BE066289 RC3-BT033
77	15	60.0	162	10	AV010258	AV010258 AV010258
78	15	60.0	169	11	BE941904	BE941904 EST421555
79	15	60.0	185	10	AL314478	AL314478 uj11h11.x
80	15	60.0	193	11	BF333453	BF333453 IL3-C1001
81	15	60.0	197	11	BF412527	BF412527 UI-R-BT1-
82	15	60.0	199	10	A1788574	A1788574 u126d08.r
83	15	60.0	200	10	BE218695	BE218695 hv43d11.x
84	15	60.0	204	10	AW627429	AW627429 hm82905.x
85	14	56.0	204	11	BE772233	BE772233 CM4-IT004

```

c 86 14 56.0 210 10 BB254937 BB254937
c 87 14 56.0 218 11 BI016166 BI016166
c 88 14 56.0 226 10 BB427689 BB427689
c 89 14 56.0 227 10 BB385773 BB385773
c 90 14 56.0 232 11 F17459 F17459
c 91 14 56.0 234 10 BB422516 BB422516
c 92 14 56.0 234 10 BB423291 BB423291
c 93 14 56.0 234 10 BB426847 BB426847
c 94 14 56.0 234 10 BB474068 BB474068
c 95 14 56.0 234 10 BB540338 BB540338
c 96 14 56.0 235 10 BB527861 BB527861
c 97 14 56.0 237 10 BB390976 BB390976
c 98 14 56.0 237 10 BB482445 BB482445
c 99 14 56.0 239 10 BB362025 BB362025
c 100 14 56.0 240 10 BB484401 BB484401

```

ALIGNMENTS

```

RESULT 1
LOCUS BF464374 324 bp mRNA EST 04-DEC-2000
DEFINITION UI-M-CG0P-bog-f-09-0-UI-s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
ACCESSION BF464374
VERSION BF464374.1 GI:11533557
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 324)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENERICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
The following repetitive elements were found in this cDNA sequence:
102-109, >(GAA)n#Simple_repeat
Seq primer: M13 Forward
POLYA-No.
FEATURES
source Location/Qualifiers
1..324
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-CG0P-bog-f-09-0-UI"
/clone_1lb="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_Ret4_S2 library is a subtracted library,
ultimately derived from mouse retina tissue libraries at
various stages of development. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_SEQ=None found"
BASE COUNT 91 a 84 c 89 g 60 t

```

```

ORIGIN
Query Match 68.0%; Score 17; DB 11; Length 324;
Best local Similarity 100.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 cgtcaggtcaggtcag 18
DB 324 Cgtcaggtcaggtcag 308

```

```

RESULT 2
LOCUS BE862745 413 bp mRNA EST 29-SEP-2000
DEFINITION UI-M-BH0-ajf-b-11-0-UI-r1 NIH_BMAP_M_S1 Mus musculus cDNA clone
ACCESSION BE862745
VERSION BE862745.1 GI:10382015
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 413)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENERICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.
FEATURES
source Location/Qualifiers
1..413
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH0-ajf-b-11-0-UI"
/clone_1lb="NIH_BMAP_M_S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_M_S1 library is a subtracted library derived from
a mixture of normalized libraries from ten regions of the
mouse brain (cerebellum, brain stems, olfactory bulbs,
hypothalamus, cortex, amygdala, basal ganglia, pineal
gland, striatum, hippocampus). The driver used for
subtraction consisted of a pool of 20,000 cDNA clones
obtained from non-normalized and normalized libraries of
these ten regions of the mouse brain."
BASE COUNT 92 a 125 c 112 g 84 t

```

```

QY 2 cgtcaggtcaggtcag 18
Query Match 68.0%; Score 17; DB 11; Length 413;
Best local Similarity 100.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


BG703880/c
 LOCUS BG703880 767 bp mRNA EST 07-MAY-2001
 DEFINITION 60268593BP1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4819446 5', mRNA sequence.
 ACCESSION BG703880
 VERSION BG703880.1 GI:13976654
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 767)
 AUTHORS NIH-MGC <http://mcc.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
 Plate: L1AM0722 row: p column: 07
 High quality sequence stop: 536.
 Location/Qualifiers
 1..767
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4819446"
 /clone_1ib="NIH_MGC_95"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcagc); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTN-3', size-selected for average insert size 2.5 kb and normalized to ROI 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 139 a 215 c 202 g 211 t
 ORIGIN
 Query Match 68.0%; Score 17; DB 11; Length 767;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 cgtcagtcgagtcag 18
 Db 649 CGTCAGGTGCGAGTCAG 633
 RESULT 6
 LOCUS AA065941/c
 DEFINITION m151a01.c1 Stralagene mouse testis (#937308) Mus musculus cDNA
 ACCESSION AA065941
 VERSION AA065941.1 GI:1562644
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 163)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
 TITLE The WashU-HMMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LINL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:309344
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 144.
 Location/Qualifiers
 1..163
 /organism="Mus musculus"
 /strain="Inbred CD-1"
 /db_xref="taxon:10090"
 /clone="IMAGE:515496"
 /clone_1ib="Stralagene mouse testis (#937308)"
 /sex="males"
 /tissue_type="testis"
 /dev_stage="10-12 week old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: testis; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'."
 BASE COUNT 23 a 56 c 49 g 35 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 10; Length 163;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 gtcagtcgagtcag 18
 Db 16 GTCAGTGCAGTCAG 1
 RESULT 7
 LOCUS BB516544/c
 DEFINITION BB516544 RIKEN full-length enriched, 16 days neonate heart Mus musculus cDNA clone D830016J11 3' similar to M76601 Mouse alpha cardiac myosin heavy chain mRNA, mRNA sequence.
 ACCESSION BB516544
 VERSION BB516544.1 GI:9568002
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 288)
 AUTHORS Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Horii,F., Ishii,Y., Ishikawa,D., Ishikawa,T., Itoh,M., Iizawa,M., Kadota,K., Kagawa,I., Kai,C., Kawaj,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomiなが,M., Toyota,T., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yamane,K., I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Kono,H., et al.)
 Unpublished (2000)

Contact: yoshihide Hayashizaki

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenomi-cho, Tsukuba, Ibaraki 305-0856, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>
 Carninci, P., Nishiyasu, Y., Westover, A., Itoh, M.,
 N. Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermostabilization and thermoactivation of thermolabile enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)
 Itoh, M., Kitsunai, T., Aizawa, J., Shibata, K., Iwawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. *Genome Res.* 9 (5), 463-470 (1999)

SOURCE

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

```
Location/Qualifiers
1..288
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="DB30016311"
/clone_id="RIKEN full-length enriched, 16 days neonatal
heart"
```

BASE COUNT	83 a	104 c	57 g	44 t
ORIGIN				

Query Match	64.0%;	Score 16;	DB 10;	Length 288;
Best Local Similarity	100.0%;	Pred. NO. 1.9e+02;		
Matches 16;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Oy	2	cgtcaggtgcaggtca	17
Db	188	CGTCAGGTGCAGGTCA	173

RESULT	8			
W54143				
LOCUS	W54143	328 bp	mRNA	EST
DEFINITION	mdj3c03.f1 Soares mouse embryo NBME13.5 14.5 Mus musculus CDNA			
DESCRIPTION	clone IMAGE:368260 5' mRNA sequence.			

ACCESSION	W54143	GI:1355214
VERSION	W54143.1	
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (bases 1 to 328)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T

Geisels, S., Kuwabara, T., Lacey, M., Ise, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterson, R.

TITLE	The Washu-HIMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@atsoni.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:229692
Seq primer: mob.REGA+ET.

Source

1.328

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clone="IMAGE:368260"
/clone_lib="Scars mouse embryo NbMEL3.5.14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"

```

/NotI="vector: pT73D-pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer 15'-
 TGTATCCATCTGTAAGTCGACGCGCCGCGAATTTTTTTTTTTTTTTTTT
 T 3',], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoura Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Fátima Bonaldo. "

```
Query Match: 64.0%; Score 16; DB 11; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	4	tcaggtgcaggtcagc	19
Db	86	TCAGGTGCAGGTCAGC	10

RESULT	9
LOCUS	W34110
DEFINITION	W34110 338 bp mRNA EST 13-MAY-1996 mc57c06.t1 Soares mouse embryo NBME13.5 14.5 Mus musculus CDNA clone IMAGE:352618 5' mRNA sequence.

VERSION	W34110.1	GI:1316081
KEYWORDS	EST.	
SOURCE	house mouse.	

ORGANISM	MUS MUSCULUS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
1 (Accession 14-220)	

REFERENCE	1 (bases 1 to 387)
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steple, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:224418
Putative full length read
Seq primer: mob.REGA+ET
High quality sequence stop: 324.

FEATURES

SOURCE

1..338

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:352618"

/clone_lib="Soares mouse embryo NBMEL3.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGAGCGCGCGCAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 j; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Felima Bonaldo."

BASE COUNT 86 a 65 c 84 g 103 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 16; DB 11; Length 338;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 tcaagtcaggtcagc 19

Db 82 TCAGGTGCGAGTCAGC 97

RESULT 10
AV633196/c 345 bp mRNA EST 15-DEC-2000

DEFINITION AV633196 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii

CDNA clone HC017h07_r 5', mRNA sequence.

AV633196

AV633196.1 GI:10776516

KEYWORDS

SOURCE

ORGANISM

Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota: Viridiplantae: Chlorophyta: Chlorophyceae: Volvocales;

Chlamydomonadaceae: Chlamydomonas.

1 (bases 1 to 345)

Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
Nakamura,Y. and Tabata,S.

Generation of expressed sequence tags from low-CO2 and high-CO2

adapted cells of Chlamydomonas reinhardtii

DNA Res. 7 (5), 305-307 (2000)

JOURNAL MEDLINE

COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp. URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

SOURCE

1..345

/organism="Chlamydomonas reinhardtii"

/strain="Cg"

/db_xref="taxon:3055"

/clone="HC017h07_r"

/clone_lib="Chlamydomonas reinhardtii 5% CO2"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"

BASE COUNT 62 a 105 c 98 g 79 t 1 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 16; DB 10; Length 345;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 caggtcaggtcagca 20

Db 173 CAGGTGCGAGTCAGCA 158

RESULT 11
A110497/c 368 bp mRNA EST 03-FEB-1997

DEFINITION m162a03.r1 Stratagene mouse testis (937308) Mus musculus CDNA

clone IMAGE:516604 5', mRNA sequence.

A110497

A110497.1 GI:1662274

EST.

KEYWORDS

SOURCE

ORGANISM

house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 368)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:310452

Seq primer: -28m13 rev1 ET from Amersham.

FEATURES

SOURCE

1..368

/organism="Mus musculus"

/strain="Inbred CD-1"

/db_xref="taxon:10090"

/clone="IMAGE:516604"

/clone_lib="Stratagene mouse testis (#937308)"

/sex="males"

/tissue_type="testis"

/dev_stage="10-12 week old"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: testis; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally; Primer:
Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor

BASE COUNT 75 a 115 c 106 g 72 t
 ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 368;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcaggtgcaggtcag 18
 |||
 Db 21 gtcaggtgcaggtcag 6

RESULT 12
 AM822106 370 bp mRNA EST 17-MAY-2000
 LOCUS

DEFINITION
 IMAGE:2802352.3', mRNA sequence.

ACCESSION
 AM822106
 VERSION
 AM822106.1 GI:7915183

KEYWORDS
 EST.

SOURCE
 house mouse.

ORGANISM
 Mus musculus

REFERENCE
 AUTHORS
 1 (bases 1 to 370)
 Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 Maria, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Peterson,
 B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritzer,
 E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R., and Wilson, R.

TITLE
 The WashU-NCI Mouse EST Project 1999
 JOURNAL
 Unpublished (1999)
 COMMENT
 Other-ESTs: uq13c09.y1
 Contact: Maria M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

FEATURES
 source
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:1041876
 Seq primer: Primer name ambiguous
 High quality sequence stop: 303.

Location/Qualifiers
 1..370
 /organism="Mus musculus"
 /strain="C3H"
 /db_xref="taxon:10090"
 /clone="IMAGE:2802352"
 /clone_lib="Ren Stubbs mouse thymus"
 /sex="mixed"
 /dev_stage="3 weeks"
 /lab_host="DH10B"
 /note="Organ: thymus; Vector: pT7T3D-Pac; Site_1: NotI;
 Site_2: PacI; 1st strand cDNA was primed with an oligo(dT)
 primer; double-stranded cDNA was ligated using 5' linker
 ggcgcgat and 3' linker aactggaagcttaatt. Library is
 size-selected >2.5 kb and average insert size is 3.5 kb.
 Clones were arrayed from primary plating; non-amplified.
 Library constructed by X. Ren and L. Stubbs (Lawrence
 Livermore National Laboratory and DOE Joint Genome
 Institute, 7000 East Ave, L-453, Livermore, CA 94550)."

BASE COUNT 125 a 57 c 75 g 113 t
 ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 370;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcaggtgcaggtcag 18
 |||
 Db 53 gtcaggtgcaggtcag 68

RESULT 13
 BF444068 395 bp mRNA EST 01-DEC-2000
 LOCUS

DEFINITION
 IMAGE:261798.MARC 2Pig Sus scrofa cDNA 5', mRNA sequence.

ACCESSION
 BF444068
 VERSION
 BF444068.1 GI:11504160

KEYWORDS
 EST.

SOURCE
 Sus scrofa

REFERENCE
 AUTHORS
 Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 1 (bases 1 to 395)
 Fahnenkrug, S.C., Fieking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
 and Keefe, J.W.

TITLE
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 JOURNAL
 Unpublished (2000)
 COMMENT
 Contact: Smith JPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

FEATURES
 source
 Email: smith@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt. trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.
 PCR primers
 FORWARD: AGGAACACGCTATGACCAT
 BACKWARD: GTTTCCAGCTCAGCAGC
 Plate: 96 row: D column: 1
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1..395
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 2Pig"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from testis, ovary,
 endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 118 a 66 c 143 g 68 t
 ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 395;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagca 20
 |||
 Db 309 CAGGTGCAGGTCAACA 324

RESULT 14
 A0298728 398 bp DNA GSS 15-DEC-1998
 LOCUS

DEFINITION
 HS_3110_B1_C12_MK CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate-3110 COL-23 Row=F, DNA sequence.

ACCESSION
 A0298728
 VERSION
 A0298728.1 GI:4015907

KEYWORDS
 GSS.

SOURCE
 human.

REFERENCE
 ORGANISM
 Homo sapiens
 Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 398)

AUTHORS Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

COMMENT 99380589

CONTACT: Mahaitas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector
Plate: 3110 row: F column: 23
Class: BAC ends

FEATURES High quality sequence stop: 398.

source Location/Qualifiers

1..398

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-3110 Col=23 Row=F"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 87 a 93 c 102 g 116 t

ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 398;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtcagtcacgacg 22
|||||

Db 330 ggtcagtcacgacg 345

RESULT 15

LOCUS A1467556/c 430 bp mRNA EST 09-MAR-1999

DEFINITION ve36909.x1 Soares mammary_gland_NbMNG Mus musculus CDNA clone

ACCESSION A1467556

VERSION A1467556.1 GI:4320893

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 430)

AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-rt@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
NCI:488568
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
High quality sequence stop: 406.

FEATURES Location/Qualifiers

1..430

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:820288"

/clone_lib="Soares_mammary_gland_NbMNG"

/sex="male"

FEATURES Location/Qualifiers

1..544

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NN0004"

/dev_stage="Adult"

/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research)

BASE COUNT 136 a 102 c 85 g 107 t

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcaggtcagtcacg 19
|||||

Db 259 TCAGGTcAGGTcAGc 244

RESULT 16

LOCUS AM892602 544 bp mRNA EST 24-MAY-2000

DEFINITION CM3-NN0004-100300-111-107 NN0004 Homo sapiens CDNA, mRNA sequence.

ACCESSION AM892602

VERSION AM892602.1 GI:8056807

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 544)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT 20202663

CONTACT: Simpson A.J.G.

LABORATORY Laboratory of Cancer Genetics

ADDRESS Ludwig Institute for Cancer Research

ADDRESS Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

TEL: +55-11-2704922

FAX: +55-11-2707001

EMAIL: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/ICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=at2-cm3-NN0004-100300-111-f07&t3=2000-03-10&t4=1>)

Seq primer: puc 18 forward

High quality sequence stop: 543.

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions.

Query Match	64.0%;	Score 16;	DB 10;	Length 544;
Best Local Similarity	100.0%;	Pred. No. 1,9e+02;		
Matches 16;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	3	gtcaagtcagtcagtcag	18	
Db	383	gtcagctcagctcag	398	

RESULT	17
A0530462/c	
LOCUS	A0530462
DEFINITION	553 bp DNA
	RPCR-11-369F11.tjt RPCR-11 Homo sapiens genomic clone RPCR-11-369F11
	, DNA sequence.
ACCESSION	A0530462
VERSION	A0530462.1
KEYWORDS	GI:4842505
SOURCE	GSS.
ORGANISM	human.
	Homo sapiens

TITLE	Use of BAC End Sequences from Library RPCT-11 for Sequence-Ready Map Building
JOURNAL	Unpublished (1997)
COMMENT	Other_GSSS: RPCT-11-369F11.TW

Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACBAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (<http://researchgenet.com>). BAC end search page: http://www.tigr.org/Tdb/huungen/bac_end_search/bac_end_search.html

Seq primer: SP6

Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .553

```

/organism="Homo sapiens"
/db_xref="GDB:7614142"
/db_xref="taxon:9606"
/clone="RPC1-11-369F11"
/clone_1ib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI
RPC11 Human Male BAC Library"
BASE COUNT      108 a      165 c      157 g      123 t

```

Query Match	Score 16	DB 13	Length 553
Best Local Similarity	100.0%	Pred. No. 1	9e+02
Matches	16	Conservative	0
		Mismatches	0
		Indels	0
		Gaps	0
Oy	3	gtcaaggtcaggtcag	18
Db	204	gttcagggtcaggtcag	189

RESULT	18
A2978539/c	
LOCUS	682 bp DNA
DEFINITION	260254K21R Mouse 10kb plasmid UNGCM library Mus musculus genomic
ACCESSION	U00024K21 R. DNA sequence.
VERSION	A2978539
KEYWORDS	A2978539.1 GI:13849766
SOURCE	GSS.
	house mouse.

TITLE	Mouse whole genome scaffolding with paired end reads from 10Kb
JOURNAL	plasmid inserts
COMMENT	Unpublished (2000)
	Contact: Robert B. Weiss

Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0254 row: K column: 21
Seq primer: CACACACGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 682.

FEATURES
SOURCE

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C2M0254K21"
/clone_lib="Mouse 10kb plasmid U06C2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/vector="PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrotomically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1pf129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

```

Query Match: Score 16; DB 13; length 682;
Best Local Similarity 100.0%; Pred. No. 1,9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY      3  gtcaggtgcagatcag 18
          |||
Db       55  gttcagggtgcaggtcag 40

```

RESULT 19
LOCUS BE535032/c
DEFINITION BE535032 775 bp mRNA EST 09-AUG-2000
601233227F1 NC1_CGAP_Mam6 Mus musculus cDNA clone IMAGE:359635 5',
mRNA sequence.
ACCESSION BE535032
VERSION BE535032
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 775)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM8775 row: f column: 08
High quality sequence stop: 225.
Location/Qualifiers
1..775
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:359635"
/clone_id="NC1_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 170 a 186 c 221 g 198 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 775;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 cgtcaggtgcaggtca 17
|||||
Db 201 CGTCAGTGCAGGTCA 186

RESULT 20
LOCUS CNS021CM
DEFINITION CNS021CM 857 bp DNA GSS 12-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
225C12 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL176719
VERSION AL176719
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 897)

AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billaud,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 897)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 897)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBA databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
source
1..897
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="225C12"
/clone_id="G"
/note="Genoscope sequence ID : CGAG225BR061P1-end : T7"
Location/Qualifiers

BASE COUNT 118 a 284 c 294 g 182 t 19 others
ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 897;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 gtcaggtgcaggtcag 18
|||||
Db 16 GTCAGGTGCAGGTGAG 31

RESULT 21
LOCUS BG172968/c
DEFINITION BG172968 976 bp mRNA EST 06-FEB-2001
602336627F1 NC1_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4459758 5',
mRNA sequence.
ACCESSION BG172968
VERSION BG172968
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 976)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10260 row: e column: 07
High quality sequence stop: 514.
Location/Qualifiers
1..976
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"

```

/clone="IMAGE:4459758"
/clone_lib="NCI_CGAP_Maml"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: Saliv;
Site 2: Noli; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT      255 a      250 c      163 g      308 t
ORIGIN

Query Match      64.0%; Score 16; DB 11; Length 976;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 gtcaggtcaggtcag 18
Db      168 gtCAGGTGCGAGTCAG 153

RESULT 22
LOCUS      AK010837      1011 bp      mRNA      HTC      05-JUL-2001
DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library,
clone:2410198H06, full insert sequence.
ACCESSION      AK010837
VERSION      AK010837.1 GI:12846558
KEYWORDS      CAP trapper.
MUS musculus (strain:C57BL/6J) ES cells cDNA to mRNA,
clone:1b:RIKEN full-length enriched mouse cDNA library
clone:2410198H06.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
1 (bases 1 to 1011)
High-efficiency full-length cDNA cloning
Methods in enzymology. 303, 19-44 (1999)
JOURNAL
MEDLINE      99279253
PUBMED      10349636
REFERENCE
AUTHORS      Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalisation and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)
JOURNAL
MEDLINE      20499374
PUBMED      11042159
REFERENCE
AUTHORS      3 (bases 1 to 1011)
Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multiplexed capillary sequencer
Genome research. 10 (11), 1757-1771 (2000)
JOURNAL
MEDLINE      20530913
PUBMED      11076861
REFERENCE
AUTHORS      4 (bases 1 to 1011)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL
MEDLINE      5 (bases 1 to 1011)
Aizawa,K., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hasegawa,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,

```

```

Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,B., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Teijima,Y., Toyota,Y., Yamamoto,T., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

FEATURES
Source      Location/Qualifiers
1..1011
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1910427"
/db_xref="MGI:1924031"
/clone="2410198H06"
/cell_type="ES cells"
/clone_lib="RIKEN full-length enriched mouse cDNA library"

BASE COUNT      255 a      227 c      180 g      349 t
ORIGIN

Query Match      64.0%; Score 16; DB 12; Length 1011;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 gtcaggtcaggtcag 18
Db      290 gTCAGGTGCGAGTCAG 275

RESULT 23
LOCUS      CNS078T3      1044 bp      DNA      GSS      08-JUL-2001
DEFINITION T7 end of clone BB0A023D11 of library BB0A from strain CBS 4732
ACCESSION      AL434413
VERSION      AL434413.1 GI:12217827
KEYWORDS      GSS.
Pichia angusta.
Pichia angusta.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE
AUTHORS      Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
and Dujon,B.
1 (bases 1 to 1044)
Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
angusta
FEBS Lett. 487 (1), 76-81 (2000)
JOURNAL
MEDLINE      20584723
PUBMED
REFERENCE
AUTHORS      2 (bases 1 to 1044)
Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bojotin-Pukhara,M., Bon,E., Brothier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durand,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neugebuse,C., Ozier-Kalogiropoulos,O., Potlery,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
JOURNAL
MEDLINE      20584711
PUBMED
REFERENCE
AUTHORS      3 (bases 1 to 1044)
Genoscope.
Direct Submission

```


Query Match 60.0%; Score 15; DB 10; Length 180;
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caagtcagtcagc 19
 |||||||||||||
 Db 138 CAGCTGCAGCTCAC 124

RESULT 26
 H22139 264 bp mRNA EST 06-JUL-1995
 LOCUS H22139/c
 DEFINITION Y138a03.s1 Soares breast 3NBHst Homo sapiens cDNA clone
 IMAGE:160492 3', mRNA sequence.

ACCESSION H22139
 VERSION H22139.1 GI:890834
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 264)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Insert Size: 685
 High quality sequence stops: 241
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 685 Std Error: 0.00
 Seq primer: Promega -21ml3
 High quality sequence stop: 241.
 Location/Qualifiers
 1..264
 /organism="Homo sapiens"
 /db_xref="GDB:574535"
 /db_xref="taxon:9606"
 /clone="IMAGE:160492"
 /clone.lib="Soares breast 3NBHst"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: breast; Vector: pT73D (Pharmacia) with a
 modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCTGAACTGGAGCGCGCCCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of a modified pT73 vector (Pharmacia).
 Library went through one round of normalization to a Cot =
 20. Library constructed by Bento Soares and M. Fatima
 Bonaldo."

BASE COUNT 87 a 47 c 50 g 73 t 7 others
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 264;
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcagtcagtcagc 18

Db 215 TCAGTGCAGTCA 201
 |||||||||||||

RESULT 27
 R50061 279 bp mRNA EST 18-MAY-1995
 LOCUS R50061/c
 DEFINITION Y159c10.s1 Soares breast 2NBHst Homo sapiens cDNA clone
 IMAGE:153042 3', mRNA sequence.

ACCESSION R50061
 VERSION R50061.1 GI:811963
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 279)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Insert Size: 681
 High quality sequence stops: 257 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 681 Std Error: 0.00
 Seq primer: Promega -21ml3
 High quality sequence stop: 257.
 Location/Qualifiers
 1..279
 /organism="Homo sapiens"
 /db_xref="GDB:565317"
 /db_xref="taxon:9606"
 /clone="IMAGE:153042"
 /clone.lib="Soares breast 2NBHst"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: breast; Vector: pT73D (Pharmacia) with a
 modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCTGAACTGGAGCGCGCCCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of a modified pT73 vector (Pharmacia).
 Library went through one round of normalization to a Cot =
 230. Library constructed by Bento Soares and M. Fatima
 Bonaldo."

BASE COUNT 93 a 49 c 54 g 75 t 8 others
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 279;
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcagtcagtcagc 18
 |||||||||||||

Db 229 TCAGTGCAGTCA 215

RESULT 28
 A1331540 292 bp mRNA EST 28-DEC-1998
 LOCUS A1331540/c
 DEFINITION fa94b01.y1 zebrafish fin day1 regeneration Danio rerio cDNA 5',

mRNA sequence.
 A1331540
 VERSION A1331540.1 GI:4068099
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.
 REFERENCE 1 (bases 1 to 292)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Knab, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
 WASHU Zebrafish EST Project 1998
 Unpublished (1998)
 TITLE JOURNAL
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@wustl.wustl.edu
 CDNA Library Preparation: Raymond Lee, CDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:
 www.rzpd.de)
 Seq primer: '3' ET from Amerham
 High quality sequence stop: 272.
 Location/Qualifiers
 1..292
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="zebrafish fln day1 regeneration"
 /sex="mixed male and female"
 /tissue_type="1 day fin regenerates"
 /lab_host="E. coli XL0LR"
 /note="Vector: pBR-CMV; Site 1: EcoRI; Site 2: XhoI; 1st
 strand CDNA primed with (CA)10ACTAGTCTCAG(T)18, followed
 by second strand synthesis, and ligated to 5' adapter (5'
)-aattgcagcag-3', 3'-gcctctc-5'. CDNA was cloned
 directionally (EcoRI/XhoI) into Stratagene Zap express
 lambda phage arms. Mass in vivo excision done to obtain
 inserts in pBR-CMV phagemid."
 BASE COUNT 55 a 83 c 96 g 58 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 10; Length 292;
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 7 gttcagtcagcagc 21
 ||||||||||||
 Db 211 gctcagtcagcagc 197
 RESULT 29
 A2877721 293 bp DNA GSS 05-MAR-2001
 LOCUS RPCI-23-195120.TV RPCI-23 Mus musculus genomic clone RPCI-23-195120
 DEFINITION 'DNA sequence.'
 ACCESSION A2877721
 VERSION A2877721.1 GI:13195787
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 293)
 Zhao, S., Niernan, W., Feldblum, T., Malek, J., Shatsman, S., Akinet,
 B., Levins, M., Moggan, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 TITLE JOURNAL
 COMMENT Other_GSSs: RPCI-23-195120.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@igf.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tlgr.org/tlgr/bac_ends/mouse/bac_end_intro.html
 Plate: 195 row: I column: 20
 Seq primer: 17
 Class: BAC ends.
 Location/Qualifiers
 1..293
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="RPCI-23-195120"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methyase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 61 a 74 c 78 g 80 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 13; Length 293;
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 gttcagtcagtcagc 17
 ||||||||||||
 Db 55 gttcagtcagtcagc 41
 RESULT 30
 BBI75346/c 300 bp mRNA EST 29-JUN-2000
 LOCUS BBI75346 RIKEN full-length enriched, adult male hypothalamus Mus
 DEFINITION musculus CDNA clone A230056E21 3', mRNA sequence.
 ACCESSION BBI75346
 VERSION BBI75346.1 GI:8834429
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 300)
 Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,
 P., Endo, F., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
 Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,

ACCESSION 3', mRNA sequence.
A1138808
VERSION A1138808.1 GI:3644780
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 382)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the L.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert Length: 593 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 362.
Location/Qualifiers
1..382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:137554"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - Oligo(dt) primer [5',
TGTTACCATCTGACGACGCGCGCCCAATTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cots5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 103 a 84 c 75 g 120 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 382;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcagca 20
|||||
Db 311 AGGTGCAGGTCAACA 325

RESULT 33
AM522699 365 bp mRNA EST 06-MAR-2000
LOCUS U1-R-B00-ahk-c-01-UI.s1 UI-R-B00 Rattus norvegicus cDNA clone
DEFINITION U1-R-B00-ahk-c-01-UI 3', mRNA sequence.
ACCESSION AM522699
VERSION AM522699.1 GI:7165124
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 385)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iuiwaeed.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized medulla library cDNA library preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLA=yes.
Location/Qualifiers
1..385
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="U1-R-B00-ahk-c-01-0-UI"
/clone_lib="UI-R-B00"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The library
(U1-R-B00) is a subtracted library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
ratel.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_L1B=U1-R-B00
TAG_L1TSUB=medulla
TAG_SEO=GAACCG"

BASE COUNT 87 a 104 c 98 g 96 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 385;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gtgcaggtcagcagc 22
|||||
Db 336 GTGCAGGTCAACG 322

RESULT 34
AM486040 396 bp mRNA EST 25-APR-2001
LOCUS AM486040
DEFINITION 69428 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AM486040
VERSION AM486040.1 GI:7056146
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 396)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Cassas,E., Wray,J.E., White,J., Cho,J., Fahnenkrug,S.C., Bennett,
G.W., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteira,G., Holt,I., Karayancheva,S., Liang,F., Quackenbush,D. and
Keeler,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred v0.960904.e. Vector identified by cross_match with the -minscore 20 and -mismatch 12 options.
PCR Primers
FORWARD: AGCAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 45 row: 1 column: 7
Seq primer: ATTACGTCACACTATAG.
Location/Qualifiers
1. .396
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 20 and day 40 embryos."

BASE COUNT 70 a 134 c 127 g 65 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 396;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 caggtgcaggtcagc 19
|||||
Db 221 CAGGTGCAGGTCAAGC 207

RESULT 35 407 bp mRNA EST 03-FEB-1997
AA062258/c LOCUS
DEFINITION m155j08.r1 StrataGene mouse testis (#937308) Mus musculus cDNA
ACCESSION AA062258
VERSION AA062258.1 GI:1556057
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 407)
AUTHORS Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marrin, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Thelning, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HIMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HIMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:309814
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 387.

FEATURES
source

Location/Qualifiers
1. .407
/organism="Mus musculus"
/strain="Inbred CD-1"
/db_xref="taxon:10090"
/clone_image="515966"
/clone_lib="Stratagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: testis; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'."
BASE COUNT 88 a 117 c 123 g 79 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 407;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 gtcaggtcaggtcagc 17
|||||
Db 15 GTCAGGTCAAGTCA 1

RESULT 36 407 bp mRNA EST 25-APR-2001
BF599674/c LOCUS
DEFINITION 263546 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF599674
VERSION BF599674.1 GI:11696393
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus

REFERENCE 1 (bases 1 to 407)
AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Cases, E., Wray, J.E., White, J., Cho, J., Fahnenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Petzel, G., Holt, L.I., Karanaycheva, S., Liang, F., Quackenbush, J. and Keeler, J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
JOURNAL 21180013
MEDLINE
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.960904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
PCR Primers
FORWARD: AGCAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 35 row: F column: 3
Seq primer: ATTACGTCACACTATAG.

FEATURES
source
Location/Qualifiers
1. .407
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal Longissimus muscle."

BASE COUNT 69 a 139 c 135 g 64 t

Query Match 60.0%; Score 15; DB 11; Length 407;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagc 19
|||||
Db 93 CAGGTGCAGGTGACG 79

RESULT 37

LOCUS A0085001/c 408 bp DNA GSS 26-AUG-1998
DEFINITION HS-2269_A2-B10_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2269 Col=20 Row=C, DNA sequence.

ACCESSION A0085001
VERSION A0085001.1 GI:3454218
KEYWORDS GSS.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector
Plate: 2269 row: C column: 20
Class: BAC ends
High quality sequence stop: 408.

FEATURES
source 1..408
Location/Qualifiers

/db_xref="taxon:9606"
/clone="Plate=2269 Col=20 Row=C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-coli DH10B"

BASE COUNT 104 a 111 c 78 g 115 t
ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 408;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagc 19
|||||
Db 396 CAGGTGCAGGTGACG 382

RESULT 38
BF653430/c 409 bp mRNA EST 25-APR-2001
LOCUS BF653430
DEFINITION 277257 MARC 3BOV Bos taurus CDNA 5', mRNA sequence.

ACCESSION BF653430
VERSION BF653430.1 GI:11918562
KEYWORDS EST.
SOURCE COW
ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Cases,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.W., Heaton,M.P., Laegreid,M.W., Rohrer,G.A., Chitko-Mckown,C.G., Percele,G., Holt,L., Karamecheva,S., Liang,F., Quakenbush,J. and Keeler,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine CDNA libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
PCR Primers
FORWARD: AGAAGACAGTATGACAT
BACKWARD: GTTCTCCAGTCAGCAGC
Plate: 69 row: P column: 10
Seq primer: ATTTAGTGACCTATAG.

FEATURES
source 1..409
Location/Qualifiers

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal Longissimus muscle."

BASE COUNT 80 a 138 c 129 g 62 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 409;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagc 19
|||||
Db 255 CAGGTGCAGGTGACG 241

RESULT 39
AA146146/c 416 bp mRNA EST 18-FEB-1997
LOCUS AA146146
DEFINITION mg65d12.r1 Soares.cthyms_2bmt Mus musculus CDNA clone IMAGE:583607
ACCESSION AA146146
VERSION AA146146.1 GI:1715582
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 416)
AUTHORS Marta,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubnue,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

TITLE
JOURNAL
COMMENT

Waterston, R.
The Washu-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:358255
Seq primer: -28M13 rev2 from Amersham.

FEATURES

source

Location/Qualifiers

1..416
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:583607"
/clone_lib="Soares_thymus_2NBMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGACAGTGGAGCGGCCGCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M. Fatima Bernaldo."

BASE COUNT

129 a 83 c 125 g 79 t

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 416;

Best Local Similarity 100.0%; Pred. No. 5.8e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 gtgcaggtcagcacg 22

Db 173 GTGCAGGTGACGACG 159

RESULT 40

AA881231/c

LOCUS 446 bp mRNA EST 26-MAR-1998

DEFINITION vx10004.r1 Soares_thymus_2NBMT Mus musculus cDNA clone

IMAGE:1264039 5', mRNA sequence.

AA881231 GI:2990541

EST.

KEYWORDS house mouse.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 446)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
Thieling, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The Washu-HHMI Mouse EST Project
Unpublished (1996)

TITLE
JOURNAL
COMMENT

Contact: Marra M/Mouse EST Project
Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

FEATURES

source

Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:665591
Seq primer: -28M13 rev2 ET from Amersham.

Location/Qualifiers

1..446
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1264039"
/clone_lib="Soares_thymus_2NBMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TTTACCAATCTGACAGTGGAGCGGCCGCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M. Fatima Bernaldo."

BASE COUNT

141 a 87 c 131 g 87 t

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 446;

Best Local Similarity 100.0%; Pred. No. 5.8e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 gtgcaggtcagcacg 22

Db 158 GTGCAGGTGACGACG 144

RESULT 41

BF604067/c

LOCUS 457 bp mRNA EST 25-APR-2001

DEFINITION BF604067 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

BF604067 GI:11702307

EST.

KEYWORDS cow.

SOURCE cow.

ORGANISM Bos taurus

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 457)
Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chikco-Wickom, C.G.,
Pettea, G., Holt, I., Karanycheva, S., Liang, F., Quackenbush, J. and
Keefe, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)

JOURNAL

MEDLINE

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68935-0166, USA
Tel: 402 762 4366
Fax: 402 762 4366

CONTACT: Smith TPJ
Email: smith@meat.ars.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT

FEATURES
source
BACKWARD: GTTTTCCGAGTCAGCAGC
Plate: 50 row: 1 column: 16
Seq primer: ATTGAGTCAGCTATAG.
Location/Qualifiers

1..457
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 380V"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPOR6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 82 a 152 c 142 g 80 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 457;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caggtgcaggtcagc 19
|||||
Db 215 CAGGTGCAGGTCA 201

RESULT 42
BI398127/c 466 bp mRNA EST 09-AUG-2001
LOCUS UNM45G05 Canine Brain cDNA Library Canis Familiaris cDNA 5', mRNA
DEFINITION sequence.
ACCESSION BI398127
VERSION BI398127.1 GI:15145202
KEYWORDS EST.
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE 1 (bases 1 to 466)
AUTHORS Roberts, M.C., Hendrickson, J.A., Hoffmann, D.E., Flickinger, G.H.,
Rutherford, M.S. and Mickelson, J.R.
TITLE University of Minnesota Canine Brain EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Mickelson, J.R.
Veterinary Pathobiology
University of Minnesota
1988 Fitch Avenue, University of Minnesota, St. Paul, MN 55108, USA
Tel: 612 625 0246
Fax: 612 625 0204
Email: mickel001@umn.edu
Seq primer: M13 Reverse.

FEATURES
source
Location/Qualifiers
1..466
/organism="Canis familiaris"
/db_xref="taxon:9615"
/clone_lib="Canine Brain cDNA Library"
/sex="Male"

/note="Organ: Brain; Vector: pSPORT1 (Gibco BRL); Site_1:
NotI; Site_2: SalI; Tissue was taken from the frontal,
occipital, temporal and parietal lobes, olfactory bulb,
hippocampus, cerebellum, thalamus, hypothalamus, midbrain
, pons, and medulla."
BASE COUNT 112 a 94 c 190 g 66 t 4 others
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 466;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 gtcaggtcaggtca 17
|||||

Db 446 GTGAGTCGAGGTCA 432

RESULT 43
AO921326/c 477 bp DNA GSS 21-DEC-1999
LOCUS RPCI-23-273p15.TV RPCI-23 Mus musculus genomic clone RPCI-23-273p15
DEFINITION , DNA sequence.
ACCESSION AO921326
VERSION AO921326.1 GI:6610329
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 477)
AUTHORS Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akınret
B., Levin, M., McGann, S., Tsengaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@delong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 273 row: p column: 15
Seq primer: T7
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..477
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-273p15"
/clone_lib="RPCI-23"
/sex="Female"

/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI methylase. size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 101 a 107 c 125 g 144 t
ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 477;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 gtcaggtcaggtca 17
|||||

Db 68 GTGAGTCGAGGTCA 54

RESULT 44
BE808281/c 492 bp mRNA EST 25-APR-2001
LOCUS BE808281 MARC 280V Bos taurus cDNA 5', mRNA sequence.
DEFINITION BE808281
ACCESSION BE808281
VERSION BE808281.1 GI:10239393
KEYWORDS EST.

SOURCE
ORGANISM
Bos taurus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Cetartiodactyla: Ruminantia: Pecora: Bovidae:
Bovidae: Bovinae: Bos.

REFERENCE
AUTHORS
1 (bases 1 to 492)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.M., Rohrer, G.A., Chitko-Mckown, C.G.,
Peters, G., Holt, L., Karaycheva, S., Liang, F., Quackenbush, J. and
Keeler, J.W.

TITLE
JOURNAL
MEDLINE
Sequence evaluation of four pooled tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013

COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mismatch 18
and mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCACGATCAGCAGC
Plate: 68 row: P column: 2
Seq primer: ATTAGTGCACACTATAG.
Location/Qualifiers
1..492
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2B0V"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPOR6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT
96 a 161 c 149 g 85 t 1 others

Query Match
Best Local Similarity 100.0%; Score 15; DB 11; Length 492;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagc 19
|||||

Db 263 CAGGTGCAGGTCAAC 249

RESULT 45
AZ261270/c
LOCUS
DEFINITION
A2261270 495 bp DNA GSS 26-JUL-2000
RPCI-23-44018.TV RPCI-23 Mus musculus genomic clone RPCI-23-44018,
DNA sequence.
ACCESSION
A2261270
VERSION
A2261270.1 GI:9469499
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
REFERENCE
1 (bases 1 to 495)
Zhao, S., Nieman, W., Felblyum, T., Malek, J., Shatsman, S., Akınret,
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
TITLE
JOURNAL
COMMENT
Mouse BAC End Sequences from Library RPCI-23
unpublished (1999)
Other_GSSs: RPCI-23-44018.TJ
Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 440 row: I column: 8
Seq primer: 77
Class: BAC ends.
FEATURES
source
1..495
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-44018"
/clone_lib="RPCI-23"
/sex="female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT
106 a 110 c 132 g 147 t

Query Match
Best Local Similarity 100.0%; Score 15; DB 13; Length 495;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcaggtgcaggtca 17
|||||

Db 66 GTCAGGTGCAGGTCA 52

Search completed: December 26, 2001, 14:10:45
JOD time: 7092 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:50:03 ; Search time 1757.07 Seconds
(without alignments) 234.726 Million cell updates/sec

Title: US-09-396-196f-2
Perfect score: 25
Sequence: 1 atgtgcgaagtcacgaattattt 25

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 220472

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 100 summaries

Database :

GenBank1: *
1: gb_ba: *
2: gb_hg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vl: *
30: em_htgo_hum: *
31: em_htgo_inv: *
32: em_htgo_rnd: *
33: em_htg_hum: *
34: em_htg_inv: *
35: em_htg_rnd: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

*

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	25	100.0	1041	6	AR029499	AR029499 Sequence
2	25	100.0	1041	6	AR034916	AR034916 Sequence
3	25	100.0	1084	6	AL1530	AL1530 B108 gene o
4	25	100.0	1121	6	E00893	E00893 Genomic DNA
5	25	100.0	5793	1	ECOBIO	J04423 E.coli 7-8-
6	25	100.0	5872	6	A38246	A38246 Sequence 1
7	25	100.0	5872	6	A38251	A38251 Sequence 6
8	25	100.0	5872	6	A93674	A93674 Sequence 1
9	25	100.0	5872	6	A93679	A93679 Sequence 6
10	25	100.0	5872	6	AR101809	AR101809 Sequence
11	25	100.0	5872	6	AR101810	AR101810 Sequence
12	25	100.0	11022	1	AE000180	AE000180 Escherich
13	25	100.0	13501	1	AE005258	AE005258 Escherich
14	25	100.0	297816	1	AF002353	AF002353 Escherich
15	24	96.0	5526	1	AF250776	AF250776 Unculture
16	17	68.0	25256	9	AP000954	AP000954 Homo sapi
17	17	68.0	32767	2	AC087305	AC087305 Homo sapi
18	17	68.0	60072	2	AC011937	AC011937 Homo sapi
19	17	68.0	86757	2	HS54178	AL133366 Homo sapi
20	17	68.0	115551	9	AL353582	AL353582 Human DNA
21	17	68.0	147706	8	AC083945	AC083945 Oryza sat
22	17	68.0	150613	2	AP003878	AP003878 Oryza sat
23	17	68.0	155172	2	AC021110	AC021110 Homo sapi
24	17	68.0	172704	2	AC092548	AC092548 Oryza sat
25	17	68.0	174815	2	AP001390	AP001390 Homo sapi
26	17	68.0	178056	2	AC091389	AC091389 Homo sapi
27	17	68.0	228516	9	AC002523	AC002523 Homo sapi
28	17	68.0	340000	9	AP001690	AP001690 Homo sapi
29	16	64.0	605	9	HSAPAF20	AF098889 Homo sapi
30	16	64.0	1829	11	G07181	G07181 human STS W
31	16	64.0	4602	9	D13644	D13644 Homo sapien
32	16	64.0	5469	1	L1A012388	AJ012388 Lactococ
33	16	64.0	43741	9	AL161800	AL161800 Human DNA
34	16	64.0	68894	2	AC074073	AC074073 Homo sapi
35	16	64.0	76075	9	AL590387	AL590387 Human DNA
36	16	64.0	105937	8	AC012679	AC012679 Arabidops
37	16	64.0	110000	9	AC044784_1	Continuation (2 of
38	16	64.0	110525	9	AB020878	AB020878 Homo sapi
39	16	64.0	111312	9	AC011248	AC011248 Homo sapi
40	16	64.0	119733	9	AL121956	AL121956 Human DNA
41	16	64.0	124103	2	AC008951	AC008951 Homo sapi
42	16	64.0	129010	9	AL159168	AL159168 Human DNA
43	16	64.0	139218	2	AC019289	AC019289 Homo sapi
44	16	64.0	140002	2	AC082893	AC082893 Homo sapi
45	16	64.0	151206	2	AC093258	AC093258 Homo sapi
46	16	64.0	153221	2	AC067998	AC067998 Homo sapi
47	16	64.0	155211	2	AC019258	AC019258 Homo sapi
48	16	64.0	156184	2	AC041018	AC041018 Homo sapi
49	16	64.0	158408	9	AL391821	AL391821 Human DNA
50	16	64.0	162383	9	AC074045	AC074045 Homo sapi
51	16	64.0	167110	2	AC022621	AC022621 Homo sapi
52	16	64.0	167803	2	AC010816	AC010816 Homo sapi
53	16	64.0	168165	9	AC073913	AC073913 Homo sapi
54	16	64.0	169671	2	AC021098	AC021098 Homo sapi
55	16	64.0	170949	2	AL590416	AL590416 Homo sapi
56	16	64.0	175931	2	AC026618	AC026618 Homo sapi
57	16	64.0	177142	2	AC026392	AC026392 Homo sapi
58	16	64.0	180251	2	AC010807	AC010807 Homo sapi
59	16	64.0	180776	2	AC032043	AC032043 Homo sapi
60	16	64.0	182125	2	AC090105	AC090105 Homo sapi
61	16	64.0	182442	2	AC013283	AC013283 Homo sapi
62	16	64.0	183166	2	AL357554	AL357554 Homo sapi
63	16	64.0	183856	2	AC021300	AC021300 Homo sapi
64	16	64.0	193970	2	AL357119	AL357119 Homo sapi
65	16	64.0	196550	9	AC020734	AC020734 Homo sapi
66	16	64.0	196975	2	AC083848	AC083848 Homo sapi
67	16	64.0	204164	9	AC020601	AC020601 Homo sapi
68	16	64.0	208656	2	AC023778	AC023778 Homo sapi
69	16	64.0	212134	2	AC024037	AC024037 Homo sapi
70	16	64.0	213510	2	AL359921	AL359921 Homo sapi

```

c 71 16 64.0 214225 2 AC026112 Homo sapi
c 72 16 64.0 216378 2 AC090420 Homo sapi
c 73 16 64.0 233454 2 AC005703 Homo sapi
c 74 15 60.0 561 6 AX136747 Homo sapi
c 75 15 60.0 1156 6 AX086780 Sequence
c 76 15 60.0 1156 6 HSM801807 Sequence
c 77 15 60.0 1771 6 AR080743 Homo sapi
c 78 15 60.0 1771 6 MOSINTESUB
c 79 15 60.0 1949 10 MOSINTESUB
c 80 15 60.0 3140 9 Y029467S18
c 81 15 60.0 3210 9 AK024286 Homo sapi
c 82 15 60.0 3520 6 AX136409 Sequence
c 83 15 60.0 3748 6 AX098224 Sequence
c 84 15 60.0 5218 9 AF152313 Homo sapi
c 85 15 60.0 5221 9 AF152315 Homo sapi
c 86 15 60.0 5233 9 AF152308 Homo sapi
c 87 15 60.0 5251 9 AF152312 Homo sapi
c 88 15 60.0 5254 9 AF152306 Homo sapi
c 89 15 60.0 5254 9 AF152310 Homo sapi
c 90 15 60.0 5257 9 AF152307 Homo sapi
c 91 15 60.0 5260 9 AF152305 Homo sapi
c 92 15 60.0 5260 9 AF152309 Homo sapi
c 93 15 60.0 5260 9 AF152311 Homo sapi
c 94 15 60.0 5260 9 AF152314 Homo sapi
c 95 15 60.0 5260 9 AF152316 Homo sapi
c 96 15 60.0 5260 9 AF152317 Homo sapi
c 97 15 60.0 5299 9 AF152303 Homo sapi
c 98 15 60.0 5431 9 AF152304 Homo sapi
c 99 15 60.0 5562 9 AB011138 Homo sapi
c 100 15 60.0 6767 9 AB051358 Homo sapi

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 attgtcgcaagtcacagaattattt 25
|||||
Db 24 ATTGTCCCAAGTCACAGATTATT 48

```

```

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE
Unknown.

```

```

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 attgtcgcaagtcacagaattattt 25
|||||
Db 24 ATTGTCCCAAGTCACAGATTATT 48

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION Bior gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Bacteriia. Proteobacteria; gamma subdivision; Enterobacteriaceae;
JOURNAL Bacteriologia.
FEATURES
source location/Qualifiers
1..1084

```

```

gene /organism="Escherichia coli"
CDS /db_xref="taxon:562"
24..1064
/gene="BIOB"
24..1064
/gene="BIOB"
/codon_start=1
/protein_id="CAA00965.1"
/db_xref="GI:490219"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 attgtcgcaagtcacagaattattt 25
|||||
Db 47 ATTGTCCCAAGTCACAGATTATT 71

```

```

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin synthetase.
E00893

```

ACCESSION E00893
 VERSION E00893.1 GI:2169154
 KEYWORDS JP 1986149091-A/1.
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1121)
 AUTHORS Hiroo, Y., Kojima, T. and Kimura, H.
 TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
 JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
 COMMENT NIPPON SODA CO LTD
 OS Escherichia coli
 PN JP 1986149091-A/1
 PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIROO YOSHITAKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
 CI2N15/00, CI2N1/20, CI2P13/18, (CI2N1/20, CI2R1.19), (CI2P13/18, PC
 CI2R1.19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli NS101;
 CC Feature is identified by experimental;
 FH Key Location/Qualifiers
 FT CDS 42..1079
 FT /product="biotin synthetase".

FEATURES
 source
 1..1121
 /organism="Escherichia coli"
 /db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t

ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 7,4e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgcgcagatcacagaattattc 25
 ||||||||||||||||||||||||
 Db 65 attgcgcagatcacagaattattc 89

RESULT 5
 ECOBIO 5793 bp DNA BCT 28-FEB-1994
 LOCUS ECOBIO
 DEFINITION E.coli 7,8-diamino-pelargonic acid (bioa), biotin synthetase (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), biotin synthetase, and dehydrobiotin synthetase (bioD), complete cds.
 J04423
 J04423.1 GI:145422
 7,8-diamino-pelargonic acid aminotransferase;
 7-keto-8-amino-pelargonic acid synthetase; bioa gene; bioB gene; bioc gene; bioc gene; bioF gene; biotin synthetase; dehydrobiotin synthetase;
 Escherichia coli (strain K-12) DNA.
 SOURCE Escherichia coli
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 1 (bases 1 to 5793)
 Otsuka, A.J., Buonocristiani, M.R., Howard, P.K., Plamm, J. and Johnson, O.
 TITLE The Escherichia coli biotin biosynthetic enzyme sequences
 J Biol. Chem. 263, 19577-19585 (1988)
 J0406784
 8906784
 Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.
 FEATURES Location/Qualifiers

source
 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 complement(98..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=11
 /protein_id="AA23513.1"
 /db_xref="GI:451568"
 /translation="MKLISNDLRDCKLPKRRHVFNGMGVGDGDNISPLHMDVDPAGTKSEVYVQYDPAATGSGMHWVNVNLPADTRVLPQGFSGSLVAMPDGYLQRTDQKKGVDGAPPKGETHRYIFETVHALDIERIDDEGASAMVGFNVHFSLSASITAMFS"
 complement(633..1925)
 /gene="bioa"
 complement(633..1925)
 /codon_start=1
 /transl_table=11
 /protein_id="AA23514.1"
 /product="7,8-diamino-pelargonic acid aminotransferase"
 /protein_id="AA23514.1"
 /db_xref="GI:457106"
 /translation="MTDDLAEFDQRHIMHPYTSMTSLPYVPVVSABCEGLILSDGRALVDGMSWMAAIIHGYNHPQLNAAKMSQIDAMSHVFGITPAELCRKLIAMSGRNALVEGVADSGVAVAEVAMKALQYQWAKGEARQFLEFRNGYHGFQDVGAMSCDDPN SMHSIMKGYLPENLFPAPAPOSRMDGEMERDVGFARLMAHRHIAVILIEPIYQVAGMKRMYHPEWMLKRIKICDRREGILLIADIELATGFRGKGLFACAEATAPDILGKALTGGITTTLSATLITREVAETISNGEAGCPHSGPTTNGRPLACAAANSLALIEGDMQDOOVADILEVQRLQAPARDAEWADVRLVAGLVETHPVMMALQFEVEQGVWRPECKLIVMPYPIILPOOLRLTAAVNAVODETFECC"
 2012..3052
 /gene="bioB"
 2012..3052
 /codon_start=1
 /transl_table=11
 /product="biotin synthetase"
 /protein_id="AA23515.1"
 /db_xref="GI:145425"
 /translation="MAHRRMTLSQVTELEFEKPLDILFEAOQVHRHFDPPQOVSTLSTKIGACPEDECKTCPTQTSRKTGLFAERLMEVQYLESAKAKAASSTPCMAANKNPHERDMPYLEDQVOGVKAMGLECMVLTGLTSESOARLANAGDLYTNHNDLTPSPYCNILITRTYORERLDLLEKVRDAGIKVCSGIVGAEYKVDAGLLIDLANPTPESVPINMLVKVGTPLADNDVDADFIRTAIVRIMKPSYVRLSAGRDQMDNQDAMCFMAGANSIFYGCKLITTPNEEDKDLQFRKLGILNPQCAVLADGNEQDQRLQALMTPDDEYYNAAL"
 3049..4203
 /gene="bioF"
 3049..4203
 /gene="bioF"
 /codon_start=1
 /transl_table=11
 /product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA23516.1"
 /db_xref="GI:145426"
 /translation="MSWQEKINAALDARRAADALRRRYPVPAQAGRMVADDDROYLNFSSMDYGLSHHPQILIRAMQOAGEFGIGSGHSGYGVVAQALEEELAEPLGYSRALIFTSQFPAANOAYIAAMAKEDRIADRSHLSLEASLSPSLRFRFHNTPILARLLASPCGQGMVTEGVSFMSDGSAPLATIQOYTOQNHMLAVDAAHICGVCGEQRCSCWLOKVPBELVTFKGFVSGAVLCSSTVADYLIOFARHLITSTSMPPAQOALRASLAIKRSDEGAREKLEALITFRAGVODLPFTLASCSAIOPLIVGDSRALQLAERLROOGCWITAIRPPYVAGTARILRLTAAHEMODIDRLLEVLHGNG"
 4190..4945
 /gene="bioc"
 4190..4945
 /gene="bioc"
 /codon_start=1
 /transl_table=11
 /protein_id="AA23517.1"
 /db_xref="GI:145427"
 /translation="MATVKNQQAIAAFGRAAAHYEOHADLQROSDALLAMLPRKYTHVIDAGCGPMGMSRHWREHRAQVTAIDLSPPMLVQARQADADHYIACDIESLPLATAHFDLAMSNLAVQWGNLSTALRELRYVVRPKGVAFPTTLVQGSLEPRHQAQWAVDERP

```

gene
    4938..5597
    /gene="b10d"
    /protein_id="AA02325.1"
    /db_xref="GI:2294845"
    /translation="MAHRRPMTLSOVTELEFEKPLDLLFEAQOVHROHPPROVOVST
    LUSIKGAPEDCKYCOSSRYKGLAEARLMEVROYLESARKAKAASSTRKGAAM
    KNHFERDMELRQMOVGKAMSLPACMTIGTISEQAOHLNAGIDYTNHLDPSPEF
    YGNITTTPTYOERLDLTKVNRDAGIKVSGGIVGLGTVKDRAGLLDLANLPPES
    VPLNMLVKKVGPLADNDVDADFTLTAVARIMMPTSYVALSAGREOMEOQOAMC
    PNAGASIFYGCKLLTPNPERDKDLQJFRKLGILNPOOTAVIAGNEDQOORLEQALMT
    PTDERYNNAAL"
    /protein_id="AA02326.1"
    /db_xref="GI:2294846"
    /translation="MATVKKOALIAAFGRAAAHYEOHADLOROSADILLMLPORKYT
    HYLDAGCGGKMSRRHREHGOVTALDLSPLMVOAROKDADADHYIAGDITSPLATA
    TPDLAMSLNAVOMCGNLSTALRELYRYVRRPGVAAFTLLVQGSIPETLROAQAVDERP
    HANRFLPDEIEOSLNGVYHOHTOPITLWFDLDSAMRSIKGIGATHLHEGRDPLIL
    TRSOLRQLQJLAMPQOOGRRPLYTHLFGVARE"
    /protein_id="AA02327.1"
    /db_xref="GI:2294847"
    /translation="MTDILAFFDORHIMHPTSMTPPYRYVVSAGCELLSDGR
    LVDCSSWMAIHGYNHNPQLAAKKSQIDAMSHVPGCITHPAIEICRKLVAWTPOP
    LECVFLDSSGVAVAVAKMALQYQAKGEAROFELFRNGYHGDFTGAMSCVDDPS
    MHSIMKGYLLENLFAPOSRMDGEMERDOWGFARLMAARHRIIAVITIEPVQNG
    GMRVHHPMLKRIKICDREGITLIDAEIATGEGRTGLPACAEIATAPILICGRKL
    TGGYMTLSATLITREVAETISNGAGCFMHPGMPGNPLCAAMASLAILLESQMOQ
    QVADIEVOLREOLAPARDAMVADVIVIGAIQVYETHPVPMALQKFFVEGQWIRP
    FGKILYMPYIILPQOLRITLAVNNAVDDEFPCQ"
    /protein_id="AA02328.1"
    /db_xref="GI:2294848"
    /translation="MKLISNDLRDGLKPHRHVFNENGSGYDGDNISPHLADWDVPAQTK
    SFVVCYDPDPAPTGSGMMHWVYVNLPADETVLPOGFGSGLVAMPDGVLTQRTDFGKTG
    YDGAAPKGETHRYLFTVHALDIERIDVDGASGAMVGFVHFHSLASASITAMFS"

RBS
    105..119
    /standard_name="PROMOTER PTAC"
    /evidence=experimental

gene
    117..1157
    /gene="BIOB"
    117..1157
    /gene="BIOB"
    /gene="BIOB"

CDS
    117..1157
    /gene="BIOB"

BASE COUNT      1363 a 1554 c 1631 g 1245 t
ORIGIN           4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match
Best Local Similarity 100.0%: Score 25; DB 1; Length 5793;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgtcgcaatgcacgaattattt 25
Db 2035 ATTGTCGCAAGTCACGATTATT 2059

RESULT 6
LOCUS      A38246      5872 bp      DNA
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION  A38246
VERSION     A38246.1  GI:2294844
KEYWORDS
ORGANISM    Escherichia coli.
            Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE       BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL    Patent: WO 9408023-A 1 14-Apr-1994;
            LONZA AG (CH)
COMMENT     other publication PL 308301 950724
            other publication CA 2345400 940414
            other publication AU 4820293 940426
            other publication HU 71781 960228
            other publication SK 42095 951108
            other publication CZ 950809 951013
            other publication FI 951547 950331
            other publication JP 85016947 960227.
FEATURES
SOURCE
    1..5872
    /organism="Escherichia coli"
    /strain="DSM498"
    /db_xref="taxon:562"
    /clone="PB030A-15/9"
    1..96
    /function="PROMOTER PTAC"
    /evidence=experimental
    23..28
    /standard_name="PROMOTER PTAC"
    /evidence=experimental
    45..49
    /standard_name="PROMOTER PTAC"
    /evidence=experimental
    105..119
    /standard_name="BIOB RBS NO.9"
    /evidence=experimental
    117..1157
    /gene="BIOB"
    117..1157
    /gene="BIOB"
    /gene="BIOB"

```

stem_loop 5583..5605
terminator 5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atctgcgaatcacagaattatt 25
|||||
Db 140 ATTGTCGCAAGTCACAGATTATT 164

RESULT 7
A38251 5872 bp DNA PAT 05-MAR-1997
LOCUS Sequence 6 from Patent WO9408023.
DEFINITION A38251
ACCESSION A38251.1 GI:2294849
VERSION
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-Apr-1994;
LONZA AG (CH)
COMMENT Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 950809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02332.1"
/db_xref="GI:2294850"
/translation="MSWQKINAAALDARRADALRRRYPAAGAGRWLVADDDQYLN
SSNDYIGISHHPOITRAMOGARPOGFGSGSGHVSQSYVHOLLEELAEWIGYSRA
LLTSGFRANQAVIAAMAKRDRFAADRLSHASLLEAASLSPSLRRFANDYTHLAR
LLASPCPGQOMVTEGVFSMDGSAPLAEIQDVTQHGWLMDADHGTGTGEOGNG
SCWLQKVPPELLVVTFGKFGVSGAAVCSSTVADYLLQFARHLIYSTMPPAQOQL
RASLAVIRSDGDAAREKLAALITFRAGVODLPTFLADSCSAIQPLIVDGNRAQL
AEKLRQGCNVTALRPPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNG"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"

CDS 3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYEYGTDEYKTVASCALLQAKAKGRTAGYKPPVASS
EKTEPGLRNSDALQRRNSQLDYPATVNPPTFAEPSPHITSQGRPESLVMSAG
LRALEQOQADWVLVEGAGGFTPLSDTFPADWVQEQLPVLLVGVKLGGINAMLA
OVIOHAGLTLGAWVANDVTPGKRHAIEYMTTLTMMIPAPLGLIPLAENPENATGK
YINLAFVDSLTGFTSR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atctgcgaatcacagaattatt 25
|||||
Db 140 ATTGTCGCAAGTCACAGATTATT 164

RESULT 8
A93674 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 1 from Patent EP0798384.
DEFINITION A93674
ACCESSION A93674.1 GI:6741862
VERSION
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-Oct-1997;
LONZA AG (CH)
COMMENT Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 950809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPRMLTSQVTELEFEKPLIDLFEAOQVHRHFDPPROVQST
LSTKGTGACPEDECKYCPQSSRYKTGLEARLMEVEQVLESARAKAAGSTRCGAAM

KNPHERMPYLEWYKAGKLEACMTLGLTSLSESQAORLANGLDYYNINLDPSPF
YGNITTTTTOEEDTLEKVDAGIKVCSGIVGIGETVDRAGLLLOLANLPTPFS
VPINMLVVKGTPLANDVDADFDIIRIIVARIIMPTSVYRLSAGREONNEOTQAC
FMAGANSIFYCKGLITTPNPEKDLQIFRLGLNPGQTVLACDNEQORLEBALMT
PDJDEYVNAAL"

RBS
/standard_name="BIOC RBS"
2284..2297
/gene="BIOC"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PINELOVY-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKNQALIAAFGRRAAHVECHADLOROSADALLMLPORKYV
HLDAGCGPMRSHRBRHAYOVALDLSPEMLVOARQKADHYLAGDIESIPLATA
TFDLAWSNLAYONGCNSTALRELYRVRKGVAFITLVGSLPELHQAWQVDEKP
HANFLEPDEILOSNGVHYOHIIQPLTWEDDLASMSRLKGTGATHLHCRDPRIL
TRSLQRLQILAMPQOGRPLVYHLPLGLVIAKE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAPFORHIMHPYTSMTSPILPYVYSAEGCELLISDRK
LVDSMSMAAHGNYNHPOLNAAMKSOIDAMSHVMGGITHAPALEICRLVAMTTP
LECFYLDSSVAVVAMKALLOYWQKCAKORFLIFRNGYHGDFFGASVCDPNS
MHSIMKCYLPENLEFAPQSRMDGEMDERMVFARLMAHREHIAVILIPYQAG
GMRVYHPWLRKIRKICDREGILLIDELATGFGRTGLFACHEAETADIIICLGAG
TGGTMTLSATLITREVAETISNGEAGFMHGPFGMGNPLACAANASIALBESGDOO
QVADIEVQLEQOLAPARDAEMVADVRLGAIIVETTHPYNMAALOKFVEQGVWIRP
EGKLITLMPYILPQOLQRLTAANRAVDDETFQCO"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDCKILPHRHFNGMGYDCNLSPLHAMDVPAGIK
SEVYTCDEDAPIGSGMHWVYVNEPADTRVLPOGGSGLVAMPDGVLOTKDFGKTC
YDGAAPKGETHYIFTHALDIERIDVDEGASGAWGPNVHFHSLASASTVAMS"
5583..5644
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgtgcgaagtcacagaattattc 25
|||||
Db 140 ATTOTGCCAGCTCAGCAATTATT 164

RESULT 9
A93679
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1.5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
1154..2308
/standard_name="BIOF RBS"
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQKINNALDARADALRRPYVAGSGRWLVADDOYLF
LSISGFRANQAVIAAMAKEDRIADRLSHASLLEASLSPSOLRRFANDVTHLAR
LLASPCPGQWVTEGVFSMDSDAPLAEIQOVTQONNGMLMWDAGTGVIGEORG
SCWLOKAPELLVYTFKGFSGVSAAYLCSSTVADYLLQFARHILYSTMPAQAOL
RASLAVIRSDGDAAREKIALITRPRAGVODLPFTLADSCSALOPLIVGDSRALQJL
AEKLRQGCWVTATRPPIVPAQTARLRLTLAHEMODIRLLEVLHGNG"
3030..3045
/standard_name="BIOF RBS"
3043..3753
/gene="BIOF"
3043..3753
/gene="BIOF"
/gene="BIOF"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="PTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYVYTGTEVGTVASCALLQAAKAGYRTAGYKPVASGS
EKTPGRLNSDALIQRNSSIOLDYAVVNPFAEPTSPHIIISAQGRPIESLVMASG
LRALEQOADMVILVAGAGGFTPLSDFTFADWVTOBOLPVILVGVKLGCTINAMLLA
QVIOHAGITLAGWANDVTPPKRHAHYMTTLRMIPAPLGLIPLAENPEANAATGK
YINLAFVASTIGTSTRL"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;

```

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattatt 25
Db 140 ATTGTGCGCAAGTCACAGATTATT 164

RESULT 10
LOCUS AR101809 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION AR101809
VERSION AR101809.1 GI:12812607
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
Source Location/Qualifiers
1..5872
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattatt 25
Db 140 ATTGTGCGCAAGTCACAGATTATT 164

RESULT 11
LOCUS AR101810 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION AR101810
VERSION AR101810.1 GI:12812608
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES
Source Location/Qualifiers
1..5872
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattatt 25
Db 140 ATTGTGCGCAAGTCACAGATTATT 164

RESULT 12
LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete
genome.
ACCESSION AE000180 U00096
```

```

VERSION AE000180.1 GI:1786988
KEYWORDS Escherichia coli K12.
SOURCE Escherichia coli K12
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 11022)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 11022)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE 3 (bases 1 to 11022)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE 4 (bases 1 to 11022)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 (e-mail: mark@ember.gatech.edu). Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.
FEATURES
Source Location/Qualifiers
1..11022
/organism="Escherichia coli K12"
/strain="K12"
/db_xref="taxon:83333"
<1..1130
/repeat_region
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
complement(147..1430)
/gene="ybhC"
/note="b0772"
```

CDS
 complement(147. .1430)
 /gene="ybhC"
 /function="orf: Not classified"
 /note="1427: 98 pct identical to fragment YBHC_ECOLI
 SW:46130 (300 aa) but contains 127 additional C-terminal
 residues"
 /codon_start=1
 /transl_table=11
 /product="putative peptidesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MTNFSVSLALALAFGVTLLACSSPPDQRFSDQAPGNSRP
 LSAEAOQNPDAOHYFASLTPGAAMNPSPTLLPACDPFVYVAGTGGTHTTQA
 AAIKRTNKROYLAVMGEYGTYYVPAAGITLTGTGKPIDVATIGLSGMSYA
 DMRDVNPGRKYMPGRKPYMYSCSRSISGICSAVFWSONGLDQLTNT
 LGSVDAGNRPAAVALRTDQVOIINNYILLGRONTFFVNSGONRLERQPTLVT
 NSTYEGVDIVSGGAVFNDTEFRVNSRTQGEAYFAFATLSNIYGLAVNSFN
 AFGGVNOLGRSLDVPANTNGQVIRDSAINEGFNPAKPMADAVISNRPACNTGSVD
 DNDEIQRLNDJTNRMMEYNNRGVSKVVEAKK"
 complement(1507. .1535)
 /note="factor Sigma70; predicted +1 start at 806574"
 complement(1582. .2058)
 /gene="ybhB"
 /note="b0773"
 complement(1582. .2058)
 /gene="ybhB"
 /function="orf: Unknown"
 /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRDGRKLRHRYFNMGYDGNISPHLWADVPA
 SEVVTYDPAPTGSGMHWVYNLPADRVLPQSGSLVAMPDGLDTRDFKTC
 YDGAAPKGETHYIFTVAHLDIRDVEGASGAVGVNFHSLASITAMES"
 complement(2108. .2124)
 /note="central position to predicted promoter: 85.5"
 /bound_moiety="Rhas predicted site"
 complement(2117. .3406)
 /gene="bioA"
 /note="b0774"
 complement(2117. .3406)
 /gene="bioA"
 /EC_number="2.6.1.62"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="1429; 100 pct identical to BIOA_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDLAFDQRIHMPYTSKMSPLPYVVASACCECILSDGRK
 LVDGSSMAALHGINHPQLNAKMSQIDAMSHVGGTTHAPALELRKIYAMP
 LECVPLADSGSAVEAMKALQWQAKGEARQRLFRNGHCTPFAMSCDPDS
 MSIMKGYLPENLFAPOSMDGEMDEDMVGFARLMAHHEIAVITPIYOGAG
 GRMYHPWLKRIRKICDEGILLADELATGFGTRGKLPCEHAEFTADLICAKL
 TGGTMTLSATLTTRVAETISNGEAGCCMGHTPGNPLACAAANASLATESGD
 OYADIEVLOLRDAPARDAEMADRYVIGATVGTTFHVNMAALQFFVEQWIKR
 PKLITYMPPTITLPQQLQRLTAAVNRAVVDTEFCQ"
 complement(2193. .2221)
 /gene="bioA"
 /note="factor Sigma70; predicted +1 start at 807260"
 3411. .3450
 /note="central position to bioB promoter: -20"
 /bound_moiety="BioB documented site"
 3411. .3450
 /note="central position to predicted promoter: -20; genetic
 evidence for the site"
 /bound_moiety="BioB predicted site"
 complement(3411. .3450)
 /note="central position to predicted promoter: 50"
 protein_bind
 protein_bind
 protein_bind
 protein_bind

protein_bind
 complement(3411. .3450)
 /note="central position to bioA promoter: 50"
 /bound_moiety="BioB documented site"
 3413. .3441
 /note="factor Sigma70; promoter bioB; documented +1 at
 808525"
 complement(3447. .3473)
 /note="factor Sigma70; promoter bioA; documented +1 at
 808515"
 3493. .4533
 /gene="bioB"
 /note="b0775"
 3493. .4533
 /gene="bioB"
 /EC_number="2.8.1.-"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0346; 99 pct identical to BIOB_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPWTLSQVTELEKPLDLLFEAQVHROHPDPROVOST
 LLSIKTGACPEDCKVYCDQSSRYKGLAEARIMEVEQYLESAKAKASTRPFCAAM
 KNPHRDMPTLEQVQVQKAMGLACMTGLTSLSQOQLRANGLDYYNNLNPSP
 YGNIITRTYGERLDTEKYDAGIKVSGVIGETVDRAGILLQLANLPPEPS
 VPIANLVKGTPLADNDVDPFDFTIVARIIMPTSYVRLSAGEQNNBTOAMC
 FMAGNSFVGCCKLITTPNPEEDKDLPKRLGINTPQOTAVLADNDDOORLQALMT
 PDPEYRNAAL"
 4530. .5684
 /gene="bioF"
 /note="b0776"
 4530. .5684
 /gene="bioF"
 /EC_number="2.3.1.47"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0384; 100 pct identical to BIOF_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSWQEKINALDARRAADALRRRYPAQAGRWLVADPROYLF
 SSNDYLGISHHPQIIRAMQGAEOFGISGSGSVHGVYHOLBEELAEMLGYRA
 LIFSGPAAOAVIAAMAKEDRIAPRAELASLSLPAASISPOLRRFANNDYTHAR
 LIASPCGOQVYVEGVSMDCSDAPLAEIQQYTOOHNGIMYDDAHGTVGEQGRG
 SCLQKVPPELLVYTFKCGVSGAAYLCSVYADYLLQFARHLIYSTMPQAQAL

Query Match 100.0%; Score 25; DB 1; Length 11022;
 Best Local Similarity 100.0%; Pred. No. 5; 9e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 attgagcaatccagattatt 25
 Db 3516 attgagcaatccagattatt 3540

RESULT 13
 AE005258
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE

AE005258 13501 bp DNA BCT 21-MAR-2001
 Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 of 135.
 AE005258 AE005174
 AE005258.1 GI:12513751
 Escherichia coli O157:H7 EDL933.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 13501)

AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., Dimantanta, E., Potamouzis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.

TITLE Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7

JOURNAL Nature 409 (6819), 529-533 (2001)

MEDLINE 21074935

PUBMED 11206551

REFERENCE 2 (bases 1 to 13501)

AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., Dimantanta, E., Potamouzis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source

1. 13501

/organism="Escherichia coli O157:H7 EDL933"

/strain="EDL933"

/serotype="O157:H7"

/db_xref="taxon:155864"

/note="enterohaemorrhagic"

misc-feature

/note="0-Island #36; Region of the EDL933 chromosome not homologous to *E. coli* K-12 MG1655; Cryptic prophage CP-933K; includes one copy of the 13 bp direct repeat that flanks the prophage"

66..665

/gene="10mk"

/note="20981"

66..665

/gene="10mk"

/function="putative membrane; Other or unknown (Phage or Prophage Related)"

/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from Genpept 118 : g117532789|gb|AA633231.1|AF151091_2 (AF151091) 10mk [prophage P-Elba]"

/codon_start=1

/transl_table=11

/product="putative outer membrane protein of prophage CP-933K"

protein_id="AAG55137.1"

/db_xref="GI:12513752"

/translation="MRKVCALILSAICLAAGVPAWASEHQSTLSAGYLHASTDPAGSDLLNGIKYKYEFTDLGLTSPSYNADEDEKTHYSDTRMHEDYVRNMFVSMAGPSVRVNMPEESAVDAVATFEXSGSGWRIDGFLVGYGF"

719..2041

/gene="20982"

719..2041

/gene="20982"

/function="putative structure; structural component (Phage or Prophage Related)"

/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from Genpept 118 : g114585436|gb|AAD25464.1|AF125520.59 (AF125520) putative tail fiber protein [Bacteriophage 933W]"

/codon_start=1

/transl_table=11

/product="putative tail component of prophage CP-933K"

protein_id="AAG55138.1"

/db_xref="GI:12513753"

/translation="MQMAVVOISGLVKDQAGKPIQNCITOLKARRNSTVAVNTVAENPDAGRYSMDEYGOVSVTLIVGGPPSHAGTISYEDSOPGTINDPLGAVTEDDARPAIRRFQWMEARHAFKRNAGCAETSRNAGTISAKASAAANADTSAPPDSESAKAKSKSEASSSASAEAKASLSLOSATDAETSKTAEASAGNARDAITSTKARSAESAQSAEORIAAEDAVNRIPITVVGPGKGPAGPGKGDGKE

gene

2489..3469

/gene="20985"

2489..3469

/gene="20985"

/function="orf; Other or unknown (Phage or Prophage Related)"

/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from Genpept 118 : g116960367|gb|AA633527.1 (AF170176) hypothetical protein predicted by Glimmer (Salmonella typhimurium LT2)"

/codon_start=1

/transl_table=11

/product="unknown protein encoded by prophage CP-933K"

protein_id="AAG55140.1"

/db_xref="GI:12513755"

/translation="MLSPITRTTHNSVNIQSSPCOTVFAGKEVLEKYDEKTPLEQWEPENPERYKQDEVPVNTKQHPYLDVNTNARIESDRNIGIVDGDSSVQKATFSKLEPDEPNWMIIVREDVDESMYDRKLSDIYHDICORLRTEDKRDYELNLELELRIEKAODSLISMTAKKRNHAMPDEFKRNALIKAGETFRCTYNTKNGHISFGEGCIYLDMDLITKIGCTIYAPDGISIMHVRNDVSNIENSATIVRSNHPALLEGISPMHSKYDABHYDOLGSGYKVFRTPLAHYNFPCDPIERNHNIIMNTSQTCSM"

3530..4522

/gene="20986"

3530..4522

/gene="20986"

/function="orf; Other or unknown (Phage or Prophage Related)"

/note="No significant matches"

/codon_start=1

/transl_table=11

/product="unknown protein encoded by prophage CP-933K"

protein_id="AAG55141.1"

/db_xref="GI:12513756"

/translation="MKIPISQSNFNSPAGSAPLAPNRAENAAVYLDIGKRILPSSAALNSVETSVIRAVHDSRSRLIDQITVDIMGTIYDALSRSGOTRDAYSITINIEKVHTICIKRYNDEYELNESSVKIDIDISLTGNELEYDVGOEPLFPCIEAGENDNEEYVSFSAVDPTSYEMPSSQGLIHEIITHVYSSDPSDSENIIEGPEITLARRAIOELGWSVPDFGVAEPRERHRLRLNALRQAAARHDEENRFAFERLGTISDYEASPDPT EYSAVSNIGYGFIOGHDFPGALINDINQDANQIOLYHGARYITFGVDKHNQ"

5350..6231

/gene="20989"

5350..6231

/gene="20989"

/function="orf; Other or unknown (Phage or Prophage Related)"

/note="No significant matches"

/codon_start=1

/transl_table=11

/product="unknown protein encoded by prophage CP-933K"

protein_id="AAG55142.1"

/db_xref="GI:12513757"

/translation="MLSPYSVNLGCSWNSLRLNLTSPDNRYLSSVRDAAVHSDNGAV

```

KVCNRYVAVATDNKFCYTRSHSGCTNLHRIKNGKCEISKIEVYLNASPVSAAM
ERGLVHSNRDLPVVDAPPPELPSVDNIRKLSVPCGNVAGNAVYEDADATKVKM
FTTSQSEVSTSEVCEENQYAGSASAKIYONNDIIGIRDKINGSLINISLPAW
AEHAIYDEEDLEOKGILFVDTETNLVYDRAKNEMFNDISSYVNSDRSMESQIMQ
SYHGKODLISVLSKI"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
/function="orf: Other or unknown (phage or prophage
related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/transl_table="MRPSTLNLVHSSRSSMSDDPDIESTVAKSSVOMTKNPKLW
OGTDNIYQOIEALDKIGSTETGRVLLNIESISRLKSTFTVILNNSRLGYMAHI
IDANHGTSDFPCNNAVEYPCGEGISVDYHATVYFELLHVRHNLNGERLAVES
SRASOKYSPILLEARTVGLGAPSEVLSSENKPFHEIGMPRTSYPSXSLIHDDNT
VSLGFQVRLHPL"
/complement(7668..8144)
/gene="yhbh"
/note="20992"
/complement(7668..8144)
/gene="yhbh"
/function="orf: Unknown function"
/residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/transl_table="MKLISNDLRDQKLPRIHRYNGMGYDGDNIISPHLAMDVPACIK
SFVTCYDPAPTGSGMMHVVNLPADPVLPGGSGIVHAPDGLOTKRDGKTC
YDGAAPKGETHYIFVHALDYERLDVDEGASGAVGVNHFHSLASITAMS"
/complement(8203..9492)
/gene="b10A"
/note="20993"
/complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 5; 8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 attgtgcaagtcacagattatt 25
|||||
Db 9602 ATTGTCGAAGTCACAGATTATT 9626
|||||

RESULT 14
LOCUS AP002553 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C., and Shingawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak

JOURNAL GENES GENET. SYST. 74 (5), 227-239 (1999)
MEDLINE 20198780
REFERENCE
2 (sites)
Makino,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C., and Shingawa,H.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
JOURNAL MEDLINE 20557356
REFERENCE
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C., and
Shingawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
JOURNAL MEDLINE 20564182
REFERENCE
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsuno,E., Nakayama,K., Murata,T.,
Tanaka,M., Toke,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M., and
Shingawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
JOURNAL MEDLINE 21156231
REFERENCE
5 (bases 1 to 297816)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shingawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/,
Fax: 81-6-6879-2047)
genome project.
COMMENT
FEATURES
location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
/note="similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/transl_table="MNLHFOAKOLPARYGLPAPVGYACTTPPEAEASKIGACPV
VKCOVHAGCRGKAGYGVNRSKEDIFAPENMLGKRLVYQDANQPNOILVEAT
DIKELITLGVNDRSSRRVEMASTGCVETEVAEETHLHKVALDPLVTGMPYOG
RELAFKLGELGKLVQOQTKTFMGLATFLERDLALIEINPLVITKQGLDLCGDKYGA
DGNALFERQDPLREMRDQSDPREAQAOQWELNVALDNGICGMVAGCLAMGTMDIV
KLHGEPANFLDVGGAATKERYTEAFKIIISDNRKAVAVNIFGIVRDLADIAGIIG
AAVEGVNVPVVRLEGNAELGAKKLADSGNLIIAKGLTDAQGVAAVAVGK"
1245..2114
/gene="ECs0754"
1245..2114
/gene="ECs0754"
/note="similar to SUCC_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1

```



```

source
1. .5526
/organism="uncultured bacterium pCosHE2"
/db_xref="taxon:143797"
/clone="pCosHE2"
/note="unknown organism, cosmid clone derived from
environmental consortium"
complement(52. .528)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kDa protein in modc-bioA
intergenic region"
/protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDGLPHRHVFNMGMDGDNISPLAMDVPAGIK
SFVVCYDPDPAITGSGMHWVYVNLPAIDTRVLFGGSGSLVAMPDGVLRTPDGNK
YDGAAPKGETHRYITFVHALDIERIDVDGASGAMGVFVHFHSLASASITAMFS"
complement(587. .1876)
/gene="bioA"
complement(587. .1876)
/gene="bioA"
/note="7,8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase BioA"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDDLAFDQRHIMPTSMT SPLPYPVVAEGCELLISDGRK
LYDMSWMAAITHGYNHPOLNAAMKSO IDAMSVHMGITTHAPALIELCKRLVMTDTP
LKCVLADSGSAEVAEMKALQYOWAKRGARPLFRNGYGLDFMGASVSDPNS
MHSIMKGYLLENFAPAPOSRMDGEMDERMVGAPFLMAHREHTAAVILIEPVQAG
GMRYHPHEMLKRIKRICDREGILLIDEITGSGRGKLPACHTAETADIDLGKAL
TGGMTSATLTTRVAVETISNCEAGCFMHPPIFGNPLACAANASLAILSGDMQO
OVADIEVQREQLAPADAEVADVYGLAGVETTHPVNMAALOKFVEGCVMTIRP
FGKLIYMPYITLLPQOLRLTAANRAVODENFEFCO"
1963. .3003
/gene="bioB"
1963. .3003
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRRMTLSQVTELEPERHLLDLEFAQCVHRRGHEDPQOVST
LSIKTGACDECKYCPQSSRYTGLFAERLMEVDVLESARKAKAAGSTRCMGAW
KNNERDMPTLEBQVGVKDLGLACMTGLTSESQAORLANAGLDVYNNHNDTPER
YGNITTRTYERLDTLEKVRDAGIKVCGGIVGLGEIYKADAGILLQANLPTPES
VPINMLVKYGTPLANDVDVDAFDFTITAVARIMPTSYVRLSAGREOMDQOAMC
FMAGANSIFYGCKLITTPNPEEDKDLQFRKLGILNPQOTAVLAGNEQOQRLQALMT
PDTDEYNAAL"
3000. .4154
/gene="bioF"
3000. .4154
/gene="bioF"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase BioF"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSWQKINAAIDARRADARRRYPAVAGAGRWLVADROYLNF
SSNDYLGISHPOIIRAMQGAOFGIGSGGSHVSGVSHQALEEELAEVLEGSKA
LFTISGFANQAVIAMAKEDRIADRLSHASLLFAASLSPSOIRREAHNIPHLAR
LIASPCGOOLVTEGVESMODSAPLAEITOOVTOHNGMLVNDASHGTGYIGBGKC
SCMIOVKRELLVYTKGFGVSGAAVICSSTVADTLLQFAHLLYSTSPDAQOAL
RASLAVTRDESDAREKLVSLIAHFRAGVQDLPETLADSCSAIOTPLVGNRSALQL
AKLRQGGCVWAIAPPIVPAGTARLRLTLTAHMODIDRLILEVLHONG"
4141. .4896
/gene="bioc"
4141. .4896
/gene="bioc"
4141. .4896
/gene="bioc"

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="B705D10"

/note="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein Bioc"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVNRQAIAAFAARALYRQHADLRQSDAVLLAMPKRYT
HYLDAGCGPQWTRHMRERAGVATLIDLSPPMLVQARQKDAADHYIAGDISLPATA
PDLIAMSNLAVQWCGNLSALBELRYVRSKGVAFPTLVGSGDPELHQAVDERP
HANRELPEIDEOSLNGVHYOHPIPTILWFEDDALSAMRSUKIGATHLHGRDPRIL
TSSQLQRQDLAMPQGGGRPLYVHLFGLVARE"
4883. .>5526
/gene="bioc"
4883. .>5526
/gene="bioc"
/note="DIB synthetase"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase Bioc"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRYFVGTDEVGKTYASCALLQAAAGRTIAGYPVAS
GSEKTPBELGRSDALQNRSSLDIDAFTVNPYFAEPTSPHIIISAGRPTESSVS
SGRLAEQOADMVLVEGAGWETPLSDFTFADWTOEOLPVLIVVGKICINHAM,
TAQALQHGALTLAGWANDVTPPKRHALEYITTLTLMIPAPLGEIPMLAENPE"

BASE COUNT      1274 a 1507 c 1567 g 1178 t

Query Match
Best Local Similarity 100.0%; Score 24; DB 1; Length 5526;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 ttgtcgcaagtcacgaattattc 25
Db      1987 TTGTCCGACGTCACGACAAATTATT 2010

RESULT 16
AP000954/c
LOCUS
DEFINITION
Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B705D10,
1156-APP region, complete sequence.
ACCESSION
AP000954
VERSION
AP000954.2 GI:6778729
KEYWORDS
HTG.
SOURCE
Homo sapiens DNA, clone:B705D10.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 25256)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 25,256 genomic DNA of 21q21.1-q21.2
Published Only in Database (1999) In press
2 (bases 1 to 25256)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-DEC-1999) to the DDBJ/EMBL/Genbank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, fax:81-42-778-9924)
On Jan 27, 2000 this sequence version replaced gi:6630654.
Sequence updated (26-Jan-2000).
FEATURES
location/Qualifiers
1. .25256
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="B705D10"

```

BASE COUNT 8452 a 4090 c 4292 g 8422 t
ORIGIN

Query Match 68.0%; Score 17; DB 9; Length 25256;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaattattt 25
|||||
Db 1714 AAGTCACAGATTATT 1698

RESULT 17
AC087305 32767 bp DNA HTG 23-DEC-2000
LOCUS Homo sapiens chromosome 18 clone RP11-238P13 map 18, LOW-PASS
DEFINITION
SEQUENCE SAMPLING.
AC087305
AC087305.1 GI:11990696
VERSION
KEYWORDS
HTG: HTGS_PHASE0.
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 32767)
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Batra, N., Bastien, V., Boguski, L., Bouhassira, B., Brown, A.,
Canales, J., Campolongo, A., Cho, K., Collins, S., Collins, S.,
Collins, S., Cooke, P., DeArnell, K., Dewar, K., Diaz, D., S.,
Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardys, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Haefford, A., Horton, L., Huie, W., Iliev, I., Johnson, R.,
Jones, C., Karas, A., Larocque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., Maclean, C., MacDonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McNeesters, R., Meltzer, J., Menus, L., Mihova, T., Mlenga, C., H.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C., H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
Soungaz, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L12071
Center clone name: 238_P_13

* NOTE: This record contains 40 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
708 807: gap of 100 bp
808 1527: contig of 720 bp in length
1528 1627: gap of 100 bp
1628 2348: contig of 721 bp in length
2349 2448: gap of 100 bp
2449 3157: contig of 709 bp in length
3158 3257: gap of 100 bp
3258 3973: contig of 716 bp in length
3974 4073: gap of 100 bp
4074 4759: contig of 726 bp in length
4800 4899: gap of 100 bp
4900 5617: contig of 718 bp in length
5618 5717: gap of 100 bp
5718 6450: contig of 733 bp in length
6451 6550: gap of 100 bp
6551 7259: contig of 709 bp in length
7260 7359: gap of 100 bp
7360 8065: contig of 707 bp in length
8067 8166: gap of 100 bp
8167 8907: contig of 741 bp in length
8908 9007: gap of 100 bp
9008 9716: contig of 709 bp in length
9717 9816: gap of 100 bp
9817 10532: contig of 716 bp in length
10533 10632: gap of 100 bp
10633 11354: contig of 722 bp in length
11355 11454: gap of 100 bp
11455 12166: contig of 712 bp in length
12167 12266: gap of 100 bp
12267 12976: contig of 710 bp in length
12977 13076: gap of 100 bp
13077 13789: contig of 713 bp in length
13790 13889: gap of 100 bp
13890 14617: contig of 728 bp in length
14618 14717: gap of 100 bp
14718 15452: contig of 735 bp in length
15453 15552: gap of 100 bp
15553 16276: contig of 724 bp in length
16277 16376: gap of 100 bp
16377 17111: contig of 735 bp in length
17112 17211: gap of 100 bp
17212 17925: contig of 714 bp in length
17926 18025: gap of 100 bp
18026 18746: contig of 721 bp in length
18747 18846: gap of 100 bp
18847 19569: contig of 723 bp in length
19570 19669: gap of 100 bp
19670 20384: contig of 715 bp in length
20385 20484: gap of 100 bp
20485 21190: contig of 706 bp in length
21191 21290: gap of 100 bp
21291 22028: contig of 738 bp in length
22029 22128: gap of 100 bp
22129 22841: contig of 713 bp in length
22842 22941: gap of 100 bp
22942 23653: contig of 712 bp in length
23654 23753: gap of 100 bp
23754 24478: contig of 725 bp in length
24479 24578: gap of 100 bp
24579 25305: contig of 727 bp in length
25306 25405: gap of 100 bp
25406 26143: contig of 738 bp in length
26144 26243: gap of 100 bp
26244 26968: contig of 725 bp in length
26969 27068: gap of 100 bp
27069 27765: contig of 717 bp in length
27766 27865: gap of 100 bp

FEATURES
source

27886 20584: contig of 699 bp in length
28585 28684: gap of 100 bp
28685 29467: contig of 783 bp in length
29468 29567: gap of 100 bp
29568 30292: contig of 725 bp in length
30293 30392: gap of 100 bp
30393 31117: contig of 725 bp in length
31118 31217: gap of 100 bp
31218 31944: contig of 727 bp in length
31945 32044: gap of 100 bp
32045 32767: contig of 723 bp in length.
Location/Qualifiers
1. 32767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-238P13"
/clone.lib="RPC1-11 Human Male BAC"

BASE COUNT 8045 a 5670 c 5396 g 9330 t 4326 others
ORIGIN

Query Match 68.0% Score 17; DB 2; Length 32767;
Best Local Similarity 100.0% Pred. No. 3.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaattat 23
|||||
Db 30730 GCAAGTCACAGAATTAT 30714

RESULT 18
AC011937
LOCUS AC011937 60072 bp DNA HNG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-16E13, LOW-PASS SEQUENCE SAMPLING.
AC011937
VERSION AC011937.2 GI:7144927
KEYWORDS HNG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 60072)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-16E13
Unpublished
2 (bases 1 to 60072)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barina, N., Beckert, R., Boguslavsky, L., Bouckghalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGuirk, A., McKernan, K., McLaughlin, J., Meltrin, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6056246.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L3522
Center clone name: 16_E_13

NOTE: This record contains 68 individual
sequencing reads that have not been assembled into
contigs. Runs of N are used to separate the reads
and the order in which they appear is completely
arbitrary. Low-pass sequence sampling is useful for
identifying clones that may be gene-rich and allows
overlap relationships among clones to be deduced.
However, it should not be assumed that this clone
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.

1 763: contig of 763 bp in length
764 863: gap of 100 bp
864 1659: contig of 796 bp in length
1660 1759: gap of 100 bp
1760 2606: contig of 847 bp in length
2607 2706: gap of 100 bp
2707 3494: contig of 788 bp in length
3495 3594: gap of 100 bp
3595 4391: contig of 797 bp in length
4392 4491: gap of 100 bp
4492 5168: contig of 677 bp in length
5169 5268: gap of 100 bp
5269 6058: contig of 790 bp in length
6059 6158: gap of 100 bp
6159 6968: contig of 810 bp in length
6969 7068: gap of 100 bp
7069 7865: contig of 797 bp in length
7866 7965: gap of 100 bp
7966 8753: contig of 788 bp in length
8754 8853: gap of 100 bp
8854 9663: contig of 810 bp in length
9664 9763: gap of 100 bp
9764 10563: contig of 800 bp in length
10564 10663: gap of 100 bp
10664 11465: contig of 802 bp in length
11466 11565: gap of 100 bp
11566 12365: contig of 800 bp in length
12366 12465: gap of 100 bp
12466 13274: contig of 809 bp in length
13275 13374: gap of 100 bp
13375 14163: contig of 789 bp in length
14164 14263: gap of 100 bp
14264 15064: contig of 801 bp in length
15065 15164: gap of 100 bp
15165 15965: contig of 801 bp in length
15966 16065: gap of 100 bp
16066 16852: contig of 787 bp in length
16853 16952: gap of 100 bp
16953 17775: contig of 823 bp in length
17776 17875: gap of 100 bp
17876 18674: contig of 799 bp in length
18675 18774: gap of 100 bp
18775 19477: contig of 703 bp in length
19478 19577: gap of 100 bp
19578 20375: contig of 798 bp in length
20376 20475: gap of 100 bp
20476 21252: contig of 777 bp in length
21253 21352: gap of 100 bp
21353 22153: contig of 801 bp in length
22154 22253: gap of 100 bp
22254 23063: contig of 810 bp in length
23064 23163: gap of 100 bp
23164 23970: contig of 807 bp in length
23971 24070: gap of 100 bp
24071 24697: contig of 627 bp in length
24698 24797: gap of 100 bp
24798 25395: contig of 598 bp in length

```

* 25396 25495: gap of 100 bp
* 25496 26303: contig of 808 bp in length
* 26304 26403: gap of 100 bp
* 26404 27193: contig of 790 bp in length
* 27194 27293: gap of 100 bp
* 27294 27964: contig of 671 bp in length
* 27965 28064: gap of 100 bp
* 28065 28724: contig of 660 bp in length
* 28725 28824: gap of 100 bp
* 28825 29615: contig of 791 bp in length
* 29616 29715: gap of 100 bp
* 29716 30499: contig of 784 bp in length
* 30500 30599: gap of 100 bp
* 30600 31391: contig of 792 bp in length
* 31392 31491: gap of 100 bp
* 31492 32268: contig of 777 bp in length
* 32269 32368: gap of 100 bp
* 32369 33177: contig of 809 bp in length
* 33178 33277: gap of 100 bp
* 33278 34070: contig of 793 bp in length
* 34071 34170: gap of 100 bp
* 34171 34978: contig of 808 bp in length
* 34979 35078: gap of 100 bp
* 35079 35870: contig of 792 bp in length
* 35871 35970: gap of 100 bp
* 35971 36751: contig of 781 bp in length
* 36752 36851: gap of 100 bp
* 36852 37644: contig of 793 bp in length
* 37645 37744: gap of 100 bp
* 37745 38573: contig of 829 bp in length
* 38574 38673: gap of 100 bp
* 38674 39462: contig of 789 bp in length
* 39463 39562: gap of 100 bp
* 39563 40367: contig of 805 bp in length
* 40368 40467: gap of 100 bp
* 40468 41259: contig of 792 bp in length
* 41260 41359: gap of 100 bp
* 41360 42171: contig of 812 bp in length
* 42172 42271: gap of 100 bp
* 42272 43084: contig of 813 bp in length
* 43085 43184: gap of 100 bp
* 43185 43981: contig of 797 bp in length
* 43982 44081: gap of 100 bp
* 44082 44891: contig of 810 bp in length
* 44892 44991: gap of 100 bp
* 44992 45788: contig of 797 bp in length
* 45789 45888: gap of 100 bp
* 45889 46671: contig of 783 bp in length
* 46672 46771: gap of 100 bp
* 46772 47570: contig of 799 bp in length
* 47571 47670: gap of 100 bp
* 47671 48460: contig of 780 bp in length
* 48461 48560: gap of 100 bp
* 48561 49352: contig of 792 bp in length
* 49353 49452: gap of 100 bp
* 49453 50256: contig of 804 bp in length
* 50257 50356: gap of 100 bp
* 50357 51147: contig of 791 bp in length
* 51148 51247: gap of 100 bp
* 51248 52076: contig of 829 bp in length
* 52077 52176: gap of 100 bp
* 52177 52976: contig of 800 bp in length
* 52977 53076: gap of 100 bp
* 53077 53891: contig of 815 bp in length
* 53892 53991: gap of 100 bp
* 53992 54778: contig of 787 bp in length
* 54779 54878: gap of 100 bp
* 54879 55683: contig of 805 bp in length
* 55684 55783: gap of 100 bp
* 55784 56484: contig of 701 bp in length
* 56485 56584: gap of 100 bp
* 56585 57369: contig of 785 bp in length
* 57370 57469: gap of 100 bp

```

```

* 57470 58276: contig of 807 bp in length
* 58277 58376: gap of 100 bp
* 58377 59169: contig of 793 bp in length
* 59170 59269: gap of 100 bp
* 59270 60072: contig of 803 bp in length.
      Location/Qualifiers
        1..60072
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone="RP11-16E13"
          /clone_lib="RPCT-11 Human Male BAC"
          /clone_1ib="RPCT-11 Human Male BAC"
BASE COUNT 15923 a 9785 c 10180 g 17174 t 7010 others
ORIGIN
Query Match 68.0%: Score 17; DB 2; Length 60072;
Best Local Similarity 100.0%: Pred. No. 3.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 9 aagtcacagatattc 25
Db 9213 AAGTCACAGATATT 9229

```

```

RESULT 19
HSS4J18/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 86757)
Nordsiek.G., Conrad.A., Dose,S., Grimm,M., Hildmann,T.,
Hornischer,K., Loehnerl.T.H., Scharfe,M., Schoen,O., Yaspo,M.L. and
Bloeker,H.
Direct Submission
Submitted (03-DEC-1999) GBF, Dept. of Genome Analysis, Mascheroder
Weg 1, D-38124 Braunschweig, Germany, E-mail: hornischer@gbf.de,
nordsiek@gbf.de, bloecker@gbf.de
On Sep 29, 2000 this sequence version replaced g1:6941908.
Collaborators:
GBF, Dept. of Genome Analysis
Mascheroder Weg 1, D-38124 Braunschweig, Germany
and
Max-Planck-Institut fuer Molekulare Genetik
Inestrasse 73, D-14195 Berlin-Dahlem, Germany
All annotations in this database entry are developed by
computational tools. It is therefore not explicitly noted in the
feature lines that evidence is not experimental.
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+++++
+ Analysis and annotation were performed with the automatic +
+ 'first-pass' annotation and submission tool +
+ 'Annotiter' (Hornischer & Bloeker). +
+ Programs used by 'Annotiter': +
+ +++++

```

```

> 'Tandem Repeats': GDE 2.2 option 'tandem'
. Minimum length 2 bp; Maximum length 20 bp; Score threshold 20 .
. Treat N's as mismatches? YES; Allow uniform consensus? NO >
> 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'Cpg
Islands': GDE 2.2 option 'cpg'
. Cpg island region size 100 bp;
. Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
e-PCR (Schuler)
. Margin: 50; Number of mismatches allowed: 0; Word size: 7 .
. STS database: 'dbSTS markers'
* NOTE: This is a 'working draft' sequence. It currently

```

```
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
13423: contig of 13423 bp in length
13424 13523: gap of 100 bp
13524 13096: contig of 1573 bp in length
15097 15196: gap of 100 bp
15197 16448: contig of 1252 bp in length
16449 16548: gap of 100 bp
16549 19018: contig of 2470 bp in length
19019 19118: gap of 100 bp
19119 24811: contig of 5693 bp in length
24812 24911: gap of 100 bp
24912 34477: contig of 9566 bp in length
34478 34577: gap of 100 bp
34578 37261: contig of 2684 bp in length
37262 37361: gap of 100 bp
37362 41884: contig of 4523 bp in length
41885 41984: gap of 100 bp
41985 45057: contig of 3073 bp in length
45058 45157: gap of 100 bp
45158 46158: contig of 1001 bp in length
46159 46258: gap of 100 bp
46259 50846: contig of 4588 bp in length
50847 50946: gap of 100 bp
50947 53362: contig of 2416 bp in length
53363 53462: gap of 100 bp
53463 55903: contig of 2441 bp in length
55904 56003: gap of 100 bp
56004 64319: contig of 8316 bp in length
64320 64419: gap of 100 bp
64420 67502: contig of 3083 bp in length
67503 67602: gap of 100 bp
67603 71620: contig of 4018 bp in length
71621 71720: gap of 100 bp
71721 74980: contig of 3260 bp in length
74981 75080: gap of 100 bp
75081 86757: contig of 11677 bp in length.
FEATURES
source
1. .86757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21"
/clone="RPC1P704J1854"
4088. .4150
/note="homology = 73.0%, counts = 7"
/rpt_family="titlata repeat"
/rpt_type=TANDEM
6744. .6807
/note="homology = 78.1%, counts = 8"
/rpt_family="titlalt repeat"
/rpt_type=TANDEM
6766. .6792
/note="TTTA repeat"
6793. .6804
/note="TAT repeat"
7177. .7192
/note="AACAA repeat"
7426. .7440
/note="AATTA repeat"
8984. .9033
/note="IR1, 92% complementary to IR1' (10757. .10806)"
/rpt_type=INVERTED
9075. .9094
/note="AAGC repeat"
10757. .10806
/note="IR1', 92% complementary to IR1 (8984. .9033)"
/rpt_type=INVERTED
satellite
12276. .12290
/note="TTG repeat"
13116. .13129
/note="ATAA repeat"
15197. .15238
/note="IR2, 97% complementary to IR2' (15480. .15521)"
/rpt_type=INVERTED
15480. .15521
/note="IR2', 97% complementary to IR2 (15197. .15238)"
/rpt_type=INVERTED
22885. .22929
/note="TG repeat"
22885. .22928
/note="homology = 100.0%, counts = 22"
/rpt_family="tg repeat"
/rpt_type=TANDEM
25803. .25870
/note="homology = 88.2%, counts = 34"
/rpt_family="at repeat"
/rpt_type=TANDEM
25819. .25843
/note="IR3, 92% complementary to IR3' (25846. .25870)"
/rpt_type=INVERTED
25828. .25870
/note="TA repeat"
25846. .25870
/note="IR3', 92% complementary to IR3 (25819. .25843)"
/rpt_type=INVERTED
26732. .26745
/note="GAA repeat"
35152. .35163
/note="TC repeat"
38666. .38678
/note="TG repeat"
40287. .40299
/note="ATT repeat"
42411. .42436
/note="IR4, 84% complementary to IR4' (44122. .44147)"
/rpt_type=INVERTED
44116. .44245
/note="homology = 63.1%, counts = 10"
/rpt_family="aaaaaaaaa repeat"
/rpt_type=TANDEM
44122. .44147
/note="IR4', 84% complementary to IR4 (42411. .42436)"
/rpt_type=INVERTED
44370. .44383
/note="AAAT repeat"
46995. .47082
/note="homology = 68.2%, counts = 8"
/rpt_family="gaaaaaaga repeat"
/rpt_type=TANDEM
47093. .47127
/note="IR5, 82% complementary to IR5' (47138. .47173)"
/rpt_type=INVERTED
47114. .47175
/note="homology = 85.5%, counts = 31"
/rpt_family="ca repeat"
/rpt_type=TANDEM
47138. .47173
/note="IR5', 82% complementary to IR5 (47093. .47127)"
/rpt_type=INVERTED
47159. .47175
/note="AT repeat"
47864. .47953
/note="homology = 66.7%, counts = 10"
/rpt_family="gaaaaaaa repeat"
/rpt_type=TANDEM
48093. .48128
/note="homology = 83.3%, counts = 18"
/rpt_family="tc repeat"
/rpt_type=TANDEM
48328. .48366
repeat_region
satellite
12276. .12290
/note="TTG repeat"
13116. .13129
/note="ATAA repeat"
15197. .15238
/note="IR2, 97% complementary to IR2' (15480. .15521)"
/rpt_type=INVERTED
15480. .15521
/note="IR2', 97% complementary to IR2 (15197. .15238)"
/rpt_type=INVERTED
22885. .22929
/note="TG repeat"
22885. .22928
/note="homology = 100.0%, counts = 22"
/rpt_family="tg repeat"
/rpt_type=TANDEM
25803. .25870
/note="homology = 88.2%, counts = 34"
/rpt_family="at repeat"
/rpt_type=TANDEM
25819. .25843
/note="IR3, 92% complementary to IR3' (25846. .25870)"
/rpt_type=INVERTED
25828. .25870
/note="TA repeat"
25846. .25870
/note="IR3', 92% complementary to IR3 (25819. .25843)"
/rpt_type=INVERTED
26732. .26745
/note="GAA repeat"
35152. .35163
/note="TC repeat"
38666. .38678
/note="TG repeat"
40287. .40299
/note="ATT repeat"
42411. .42436
/note="IR4, 84% complementary to IR4' (44122. .44147)"
/rpt_type=INVERTED
44116. .44245
/note="homology = 63.1%, counts = 10"
/rpt_family="aaaaaaaaa repeat"
/rpt_type=TANDEM
44122. .44147
/note="IR4', 84% complementary to IR4 (42411. .42436)"
/rpt_type=INVERTED
44370. .44383
/note="AAAT repeat"
46995. .47082
/note="homology = 68.2%, counts = 8"
/rpt_family="gaaaaaaga repeat"
/rpt_type=TANDEM
47093. .47127
/note="IR5, 82% complementary to IR5' (47138. .47173)"
/rpt_type=INVERTED
47114. .47175
/note="homology = 85.5%, counts = 31"
/rpt_family="ca repeat"
/rpt_type=TANDEM
47138. .47173
/note="IR5', 82% complementary to IR5 (47093. .47127)"
/rpt_type=INVERTED
47159. .47175
/note="AT repeat"
47864. .47953
/note="homology = 66.7%, counts = 10"
/rpt_family="gaaaaaaa repeat"
/rpt_type=TANDEM
48093. .48128
/note="homology = 83.3%, counts = 18"
/rpt_family="tc repeat"
/rpt_type=TANDEM
48328. .48366
repeat_region
```



```

misc_feature complement(18532..18952)
      /note="match: GSS: Em:A0633379"
repeat_region /note="FLAM_A repeat: matches 1..133 of consensus"
misc_feature /note="match: GSS: Em:A0122529"
      19472..19617
repeat_region /note="MIR repeat: matches 78..218 of consensus"
      20285..20513
repeat_region /note="MTLTD repeat: matches 320..501 of consensus"
      20695..20832
repeat_region /note="MIR repeat: matches 53..189 of consensus"
      20924..21387
repeat_region /note="MTLTD repeat: matches 2..505 of consensus"
      21818..21867
repeat_region /note="MER9A repeat: matches 1..48 of consensus"
      22143..22304
repeat_region /note="MIR repeat: matches 31..192 of consensus"
      22339..22584
repeat_region /note="MTLTD repeat: matches 1..225 of consensus"
      22585..23315
repeat_region /note="LTR39 repeat: matches 1..794 of consensus"
      23316..23455
repeat_region /note="MTLTD repeat: matches 225..370 of consensus"
      24413..24566
repeat_region /note="MIR repeat: matches 29..200 of consensus"
      24745..25217
misc_feature /note="match: GSS: Em:A0403530"
      24758..25162
misc_feature /note="match: GSS: Em:A0435639"
      25811..25956
repeat_region /note="MIR repeat: matches 54..232 of consensus"
      27482..27631
repeat_region /note="5 copies 30 mer 77% conserved"
      27497..27631
repeat_region /note="9 copies 15 mer 79% conserved"
      27635
misc_feature /note="Tandem repeat. Forced join. Gap size estimated to
      be approximately 320bp by HindIII, SacI and SspI
      restriction enzyme digest data."
      28011..28380
repeat_region /note="MTB repeat: matches 1..426 of consensus"
      29147..29602
repeat_region /note="LTR24 repeat: matches 1..490 of consensus"
      30051..30224
repeat_region /note="2 copies 87 mer 92% conserved"
      32247..32977
repeat_region /note="LTPA16 repeat: matches 5423..6154 of consensus"
      33453..33754
repeat_region /note="LTR8 repeat: matches 390..691 of consensus"
      34281..34449
repeat_region /note="BURL repeat: matches 8180..8350 of consensus"
      34808..34911
repeat_region /note="4 copies 26 mer 78% conserved"
      34811..34902
repeat_region /note="46 copies 2 mer 78% conserved"
      34813..34902
repeat_region /note="3 copies 30 mer 84% conserved"
      34821..34910
repeat_region /note="9 copies 10 mer 74% conserved"
      34914..34951
repeat_region /note="3 copies 26 mer 79% conserved"
      34916..34975
repeat_region /note="6 copies 10 mer 74% conserved"
      35201..35320
repeat_region /note="MER9A repeat: matches 1..123 of consensus"
      35497..35835
repeat_region /note="LTPA8 repeat: matches 5819..6162 of consensus"
      35836..36095
repeat_region /note="Alu repeat: matches 39..304 of consensus"
      36096..40074
repeat_region /note="LTPA8 repeat: matches 1897..5819 of consensus"
      40075..40337

```

```

Query Match      68.0%; Score 17; DB 9; Length 11551;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 aagtcacgaatcatt 25
Db 72877 AGCTCACAGATATATTT 72861

RESULT 21
AC083945/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 147706)
Spiegel, L.A., King, L., Kirchoff, K.A., de la Bastide, M.,
Preston, R.R., Nascimento, L.U., Vill, M.D., Baker, J.P., Miller, B.,
Cunnius, D.M., Kuit, K.H., Rodriguez, S., Santos, L., Zuluvern, T.,
Ballja, V.S., Shah, R.S., Bahret, A., Bal, H.P., O'Shaughnessy, A.,
Dedhia, N.N. and McCombie, W.R.
Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
Clone OSJNBa0058E19, Complete Sequence
Unpublished
2 (bases 1 to 147706)
McCombie, W.R.
Direct Submission
Submitted (07-OCT-2000) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
NY 11724, USA
3 (bases 1 to 147706)
McCombie, W.R.
Direct Submission
Submitted (10-JAN-2001) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
NY 11724, USA
4 (bases 1 to 147706)
Bal, H.P., Spiegel, L.A., King, L., Kirchoff, K.A., de la Bastide, M.,
Preston, R.R., Nascimento, L.U., Vill, M.D., Baker, J.P., Miller, B.,
Cunnius, D.M., Kuit, K.H., Rodriguez, S., Santos, L., Zuluvern, T.,
Ballja, V.S., Shah, R.S., Bahret, A., O'Shaughnessy, A., Dedhia, N.N.
and McCombie, W.R.
Direct Submission
Submitted (27-FEB-2001) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor, NY
11743, USA
5 (bases 1 to 147706)
Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
Clone OSJNBa0058E19, Complete Sequence.
Bal, H.P., Spiegel, L.A., King, L., Kirchoff, K.A., de la Bastide, M.,
Preston, R.R., Nascimento, L.U., Vill, M.D., Baker, J.P., Miller, B.,
Cunnius, D.M., Kuit, K.H., Rodriguez, S., Santos, L., Zuluvern, T.,
Ballja, V.S., Shah, R.S., Bahret, A., O'Shaughnessy, A., Dedhia, N.N.
and McCombie, W.R.
Direct Submission
Submitted (08-MAR-2001) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor, NY
11743, USA
REMARK
COMMENT
Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
Clone OSJNBa0058E19, Complete Sequence.
On Jan 10, 2001 this sequence version replaced gi:11038452.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such

```

as compressions and repeats: all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

FEATURES

SOURCE

```

1..147706
/organism="Oryza sativa"
/db_xref="taxon:4530"
/chromosome="X"
/clone="OSJNBa0058E19"
/clone_1fb="HindIII"
58..5309
/gene="OSJNBa0058E19.1"
/note="Histidine kinase-like protein"
join(58..1156,455..1156,1654..2157,2536..2693,2853..2957,
4891..5136,5200..5309)
/gene="OSJNBa0058E19.1"
/codon_start=1
/product="Histidine-kinase like protein"
/protein_id="AAK13126.1"
/db_xref="GI:13129468"
/translaton="MAVLVSDQVDPVLIJGDPWFRRITITNVCNSKKEFTBOGHIFIRV
HLIEVYRKMEALDPTSPENIEVTANSKNTMPYNTLSGLEVANNRRTLSERFKKOS
DAIDSVMLVTVEEDTIGITKDAQTRIFTPPMQADGSTSKTYGTGIGISTIKREL
MGEGIGVSKRGVSTSPFAIKENKDKDGRKRCPEPTPDPOGMALVYDGRCA
RAVYTMHLRRLIGCDLATSALIESCNSSVSKSLNVLVDKLEAMGDSGLA
EPFLCIDLRKLTLSKQMPKPFLLAGSTTPADSPCLRLAGVNSIRKPLRSTVAA
CLSKALCVLTGRSRDNLVLSVLTGRNLIYVDNANRIVAAALKKYGAIVYCV
DSKEATSLRQPHKEDPACHVQMPDMGFEATRLVRSVESKINDTIOAGEVSELY
GNMAHVPILAMTADVIQATPEGSASKNPFQGNIAKOSYAVNREOKDWDVAIYS
FEPTHPDSTFNTGRAGPDGRAVPOIAGAPHPRSLVAAVQDTGANMNCN
DTEKLTQTRSQFLPMRRCRRLALHDSRVVVVGGGGSVCAVTPMSSAIDCN
RIC"
4968..7539
/note="Micropon4 Oryza sativa gene, repeat sequence
Micropon"
complement(7197..10518)
/gene="OSJNBa0058E19.2"
/note="Putative protein kinase Xa21"
complement(join(7197..7319,7661..7668,7886..8184,
8253..8972,9073..9099,9272..9425,10121..10354,
10369..10518))
/gene="OSJNBa0058E19.2"
/codon_start=1
/product="Putative protein kinase Xa21"
/protein_id="AAK13127.1"
/db_xref="GI:13129469"
/translaton="MNLKLSLALSYLLFLAANAATAATRCETIEIVKTTGDRNAGTD
ARVSLQCPDADRGQGVGPDGRRLRREGEPVYGRRLHAGVLAHQAHRLRM
EOARVRELRLDGDARAGETPVDAPVANGPKNFGMLRRSHDPTTYVWVGKRPV
POLYIMOKLPTKAPKPTGGEVDSGSAASAASAGRCSTPTSHKARVRSQ
CAPSPGSGSSRAVQILRVSRGALVYVRAAEASSTPTPEPOTTRAAAVASPL
PALNLTNGLIGQSPNLNLTFLKFLVNPANSTFGLIPQSGNHHILQITLYLSNLTIO
RRMPLNANCSNKLKVLNGLNGLVGOIPADLPQRFQSLLSVNSITGPIIPSVAKITTL
KRFGCLYNIDIGSIPDNFAKLPGLVYLHLGAKIIGSGECLPNSIGNSLPDQKPOLGN
FEYGHINPNSLTNASKNLIDIVINSFTGVPVRSYTGKLTLSMNLINLEINHSQCD
LEFNSLANCTELOMSIYGVGVELAPSAARYKSHPRENPVSAKPKPGPSSSPN
PAKARSRSAIAPGATLVYTRLPAAAAPEDSGAPAVRGELDQDGAAGDDAAAAAT
HLAERTPAORSREKREKGRNKRKSTVWHPDMGPHGSHDSKAM"
7923..9998
/note="Retrosat2 Ty3-Gypsy-retroelement"
18596..19460
/gene="OSJNBa0058E19.3"
/note="Hypothetical protein"
join(18596..18726,18794..18896,18917..18991,19260..19460)
/gene="OSJNBa0058E19.3"
/codon_start=1
/product="Hypothetical protein"
/protein_id="AAK13128.1"
/db_xref="GI:13129470"
/translaton="MNLIVQIRKMSIFLLHSTISSKAAGLWPSARCRROWTPIC
GHRSSASGESEDRMGSGGCACRAIDWMTPKCMCHGTSSDEVGVGLSDSDADVPS
ALYLIHSNRNRRLDLVAVHVCVSGAPAGEVAFPPNHCMTAEELRGDTGTERRMWT
REKTIAMQ"

```

gene	complement(22790..40376)	gene	complement(22790..40376)
CDS	/gene="OSJNBa0058E19.4"	CDS	/gene="OSJNBa0058E19.4"
	/note="Putative gag-pol polypeptide"		/note="Putative gag-pol polypeptide"
	complement(join(22790..23208,24472..24598,24918..25000,		complement(join(22790..23208,24472..24598,24918..25000,
	25046..25158,26244..26431,26674..26955,27044..27126,		25046..25158,26244..26431,26674..26955,27044..27126,
	28396..28467,28767..28900,29535..29647,30186..30251,		28396..28467,28767..28900,29535..29647,30186..30251,
	30344..30686,31565..31711,32362..32410,33038..33286,		30344..30686,31565..31711,32362..32410,33038..33286,
	34586..34714,35757..36060,36212..36315,37016..37109,		34586..34714,35757..36060,36212..36315,37016..37109,
	37580..37612,37836..37921,38120..38188,39131..39252,		37580..37612,37836..37921,38120..38188,39131..39252,
	39319..39444,39499..40376)		39319..39444,39499..40376)
	/gene="OSJNBa0058E19.4"		/gene="OSJNBa0058E19.4"
	/codon_start=1		/codon_start=1
	/product="Putative gag-pol polypeptide"		/product="Putative gag-pol polypeptide"
	/protein_id="AAK13129.1"		/protein_id="AAK13129.1"
	/db_xref="GI:13129471"		/db_xref="GI:13129471"
	/translaton="MDSITMSNGTWEVRRPCKPVGCKVFKKKLRPDGTIEYKAR		/translaton="MDSITMSNGTWEVRRPCKPVGCKVFKKKLRPDGTIEYKAR
	LVAKGYTQKEGEDEFDITYSPVRLTTITVLLALASHLIHLQMDVTFALINLEEE		LVAKGYTQKEGEDEFDITYSPVRLTTITVLLALASHLIHLQMDVTFALINLEEE
	ILMDPDGVVLESQGMWCKLKLKSLYGLKQAPKQMKHEKFDNTLSAGVYNEADKCY		ILMDPDGVVLESQGMWCKLKLKSLYGLKQAPKQMKHEKFDNTLSAGVYNEADKCY
	YRCGGEVILCLVYDILIFGTSLVNIEEYKDELSKFEMTDLRVADVLINILKLG		YRCGGEVILCLVYDILIFGTSLVNIEEYKDELSKFEMTDLRVADVLINILKLG
	DEGGLTVOSHVQKVLSPFGSDCKAPPTPDPSVLRLKRRRLAROLRYSOITIGSL		DEGGLTVOSHVQKVLSPFGSDCKAPPTPDPSVLRLKRRRLAROLRYSOITIGSL
	MYLASFATPDISEAFSKMYRLKGTMSYIYAGYKPLKEGSDSNMTSDSEIKANS		MYLASFATPDISEAFSKMYRLKGTMSYIYAGYKPLKEGSDSNMTSDSEIKANS
	GSTMREILTDALDVAEMLRELLMDLPVYKRPVALINMGACRGACRGTATFAL		GSTMREILTDALDVAEMLRELLMDLPVYKRPVALINMGACRGACRGTATFAL
	SEFVDAITRTEEDVILHDSCTSSSSVVELQNPVLVGTSCRLNSRLRTCTRRAL		SEFVDAITRTEEDVILHDSCTSSSSVVELQNPVLVGTSCRLNSRLRTCTRRAL
	KELGASASATVAVLDYCYCIDAFAVNSMETIGDEKNDINGNTRASSSGGTFLGNL		KELGASASATVAVLDYCYCIDAFAVNSMETIGDEKNDINGNTRASSSGGTFLGNL
	REILISGRKFTVANSMPSTYKLDVLEFEFEKHYVFGMEKAKKAPGLDGFPAEL		REILISGRKFTVANSMPSTYKLDVLEFEFEKHYVFGMEKAKKAPGLDGFPAEL
	EHKGNFVETKMANRILVAKQVIRPQTAFAFKRNIMEGAIIHETLHDMCKKSD		EHKGNFVETKMANRILVAKQVIRPQTAFAFKRNIMEGAIIHETLHDMCKKSD
	GVILKLYFEKSYDKENMNLQOALRMKSEYIMFKWIDVSGSVAVPKYICGPDVY		GVILKLYFEKSYDKENMNLQOALRMKSEYIMFKWIDVSGSVAVPKYICGPDVY
	LGPGLTLYVLCSTLILKRGVETKTYILIREKAPVAVCSAASASARAPSLPRVPSA		LGPGLTLYVLCSTLILKRGVETKTYILIREKAPVAVCSAASASARAPSLPRVPSA
	AAAPVAVABRHRQVAVATVILSSYSKACRBAALIDPRTAASTSLPSSKRP		AAAPVAVABRHRQVAVATVILSSYSKACRBAALIDPRTAASTSLPSSKRP
	SESLICILACRRREGDYPGKRCVELSLVRLVTTQMTVCSSAAHDMWYTCG		SESLICILACRRREGDYPGKRCVELSLVRLVTTQMTVCSSAAHDMWYTCG
	OVSCGYSPPTRELSEKRCGODASVSLTGGLADLRRCAPAPNLAKVENEHMYRFR		OVSCGYSPPTRELSEKRCGODASVSLTGGLADLRRCAPAPNLAKVENEHMYRFR
	GVSGHPAHNAIRGEHTSEDPTSDAQBEGVYRILRLGFLQDCEWBPASDDKADKT		GVSGHPAHNAIRGEHTSEDPTSDAQBEGVYRILRLGFLQDCEWBPASDDKADKT
	RLRVOLCKWPLPLYLDLGTNTNTNGVSLNMSMTGMDGQDVDRITACVRRRAAR		RLRVOLCKWPLPLYLDLGTNTNTNGVSLNMSMTGMDGQDVDRITACVRRRAAR
	HTPSNNDGSFEERWDRKRRRANGAEGSAVGLNRRHROGARPASSPDGAAGORLVLP		HTPSNNDGSFEERWDRKRRRANGAEGSAVGLNRRHROGARPASSPDGAAGORLVLP
	LAGYCARPOSTNRPKLOEISRPVAKKLPSPQGLNREBSGKRLASDTOTTERET		LAGYCARPOSTNRPKLOEISRPVAKKLPSPQGLNREBSGKRLASDTOTTERET
	SCITPGATVGNMGPGVPTVPLHMPKRSRTAFVNVGSLNMTSGRTTDTGTDVRR		SCITPGATVGNMGPGVPTVPLHMPKRSRTAFVNVGSLNMTSGRTTDTGTDVRR
	IACVRRRAARHIPSNNDEGRSIVGRNLEAWELGKSVNVALTDQKRPFWKNNKDG		IACVRRRAARHIPSNNDEGRSIVGRNLEAWELGKSVNVALTDQKRPFWKNNKDG
	IVPRKFEIKLKLPLKISFLWLYFEMGTIARVMPILKGVDEWVSKYCKL		IVPRKFEIKLKLPLKISFLWLYFEMGTIARVMPILKGVDEWVSKYCKL
	EKRALEFEMVYVTTMHSIEAFHPRAPSMHASRIVMTYKOHQDAEGCEVYKKEEPS		EKRALEFEMVYVTTMHSIEAFHPRAPSMHASRIVMTYKOHQDAEGCEVYKKEEPS
	MCCGCKRTKPKARILLSTIOSORRKKVFLVMPITRSCEDHMSAEVAVMGRCAFAV		MCCGCKRTKPKARILLSTIOSORRKKVFLVMPITRSCEDHMSAEVAVMGRCAFAV
	LGRAGFPYGEKMMAMAFKRIIDATCSTGCVGTG"		LGRAGFPYGEKMMAMAFKRIIDATCSTGCVGTG"
	25681..29265		25681..29265
	/note="retro02 retroelement"		/note="retro02 retroelement"
	30888..31407		30888..31407
	/note="RIIE4 Oryza sativa gypsy-type retrotransposon"		/note="RIIE4 Oryza sativa gypsy-type retrotransposon"
	complement(30963..32572)		complement(30963..32572)
	/note="RIIE8 Oryza sativa gypsy-type retrotransposon"		/note="RIIE8 Oryza sativa gypsy-type retrotransposon"
	31529..31692		31529..31692
	/note="Truncator transposon from Oryza sativa"		/note="Truncator transposon from Oryza sativa"
	complement(31804..32083)		complement(31804..32083)
	/note="RIIE8 Oryza sativa gypsy-type retrotransposon"		/note="RIIE8 Oryza sativa gypsy-type retrotransposon"
	32562..34237		32562..34237
	/note="Truncator transposon"		/note="Truncator transposon"
	complement(32638..33296)		complement(32638..33296)
	/note="RIIE8 Oryza sativa gypsy-type retrotransposon"		/note="RIIE8 Oryza sativa gypsy-type retrotransposon"
	complement(34475..43713)		complement(34475..43713)
	/gene="OSJNBa0058E19.5"		/gene="OSJNBa0058E19.5"
	/codon_start=1		/codon_start=1
	/product="Putative gag-pol polypeptide"		/product="Putative gag-pol polypeptide"
	/protein_id="AAK13130.1"		/protein_id="AAK13130.1"
	/db_xref="GI:13129472"		/db_xref="GI:13129472"

Query Match 68.0%: Score 17; DB 8; Length 147706;
 Best Local Similarity 100.0%: Pred. No. 3.2;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 8 caagtcacagaataatt 24
 ||||||||||||||||

DB 10686 CAAGTCACAGATTATT 10670

RESULT 22
AP003878
LOCUS
DEFINITION Oryza sativa chromosome 8 clone OJ1120.C08, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AP003878
VERSION AP003878.1 GI:14646811
KEYWORDS HTG; HTGS-PHASE2.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1120.C08.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretidae; Oryzaceae; Oryza.
REFERENCE
AUTHORS 1 (bases 1 to 150613)
TITLE Sasaki,T., Matsumoto,T. and Yamamoto,K.
JOURNAL Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC
clone:OJ1120.C08
Published Only in Database (2001) In press
2 (bases 1 to 150613)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (09-JUL-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:sasaki@abrc.affrc.go.jp, URL:http://gyp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps is believed
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Source
1..150613
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="8"
/clone="OJ1120.C08"

BASE COUNT 45837 a 30197 c 29818 g 44602 t 159 others

ORIGIN

Query Match 68.0%; Score 17; DB 2; Length 150613;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacagattatt 24
|||||
Db 72720 CAAGTCACAGATTATT 72736

RESULT 23
AC021110
LOCUS
DEFINITION AC021110 155172 bp DNA HTG 07-JUL-2000
Homo sapiens chromosome 18 clone RP11-238P13, WORKING DRAFT
SEQUENCE, 6 unordered pieces.
AC021110
VERSION AC021110.3 GI:8439978
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 155172)
TITLE Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
unpublished
2 (bases 1 to 155172)
Waterston,R.H.
Direct Submission
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 10, 2000 this sequence version replaced gi:7021826.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
Center project name: H_NH0238P13
----- Project Information -----
Sequencing vector: M13: 73%
Sequencing vector: plasmid: 27%
Chemistry: Dye-terminator Big Dye, 27% of reads
Chemistry: Dye-terminator Big Dye, 27% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151162 bases at least Q40
Consensus quality: 152443 bases at least Q30
Insert size: 177000; agarose-fp
Insert size: 154672; sum-of-contigs
Quality coverage: 4.08 in Q20 bases; agarose-fp
Quality coverage: 4.69 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 9265: contig of 9265 bp in length
* 9366: gap of unknown length
* 9366: contig of 17139 bp in length
* 26505: gap of unknown length
* 26505: contig of 17696 bp in length
* 44301: gap of unknown length
* 44301: contig of 17850 bp in length
* 62251: gap of unknown length
* 62251: contig of 43394 bp in length
* 105745: gap of unknown length
* 105745: contig of 49328 bp in length
* 105845: 155172: contig of 49328 bp in length.

FEATURES
Source
1..155172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/clone="RP11-238P13"
1..9265
/note="assembly_name:Contig1"
9366..26504
/note="assembly_name:Contig2"
26505..44300
/note="assembly_name:Contig3
clone_end:47
vector_side:right"
44401..62250
/note="assembly_name:Contig4"
62351..105744
/note="assembly_name:Contig5"
105845..155172
/note="assembly_name:Contig6"

BASE COUNT 51543 a 28385 c 28043 g 46698 t 503 others

ORIGIN

Query Match 68.0%; Score 17; DB 2; Length 151172;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaattatc 23
 |||
 Db 111929 GCAAGTCACAGAAATTAT 111945

RESULT 24
 AC092548/c
 LOCUS
 DEFINITION
 AC092548.2
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AC092548 172704 bp DNA HTG 14-AUG-2001
 Oryza sativa chromosome 10 clone OSJNBa0073L01, *** SEQUENCING IN
 PROGRESS *** 10 unordered pieces.
 AC092548.2 GI:15150615
 HTG: HTGS_PHASE1.
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; Oryza.

REFERENCE
 AUTHORS
 TITLE

1 (bases 1 to 172704)
 Buell, R., Hsiao, J., Zismann, V., Motfal, K.M., Hill, J.,
 Gansberger, K., Burgess, S., Jarrahi, B., Shvartsbeyn, M., Brenner, M.,
 Clecko, A., Pal, G., Vanaken, S., Hansen, C., Ullrich, T.,
 Feldblyum, T., Khalak, H.G., Yuan, Q., Quackenbush, J., White, O.,
 Salzberg, S. and Fraser, C.
 Oryza sativa ssp. japonica cv. Nipponbare OSJNBa0073L01 BAC genomic
 sequence
 Unpublished
 2 (bases 1 to 172704)
 Buell, R.

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Direct Submission
 Submitted (17-JUL-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 On Aug 14, 2001 this sequence version replaced gi:14787122.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4727: contig of 4727 bp in length
 * 4728 4772: gap of unknown length
 * 4773 4773: gap of unknown length
 * 20213 20257: gap of unknown length
 * 20258 87269: contig of 67012 bp in length
 * 87270 87314: gap of unknown length
 * 87315 96752: contig of 9438 bp in length
 * 96753 96797: gap of unknown length
 * 96798 106410: contig of 9613 bp in length
 * 106411 106454: gap of unknown length
 * 106455 117131: contig of 10677 bp in length
 * 117132 117175: gap of unknown length
 * 117176 129369: contig of 12194 bp in length
 * 129370 129413: gap of unknown length
 * 129414 143358: contig of 13945 bp in length
 * 143359 143402: gap of unknown length
 * 143403 157616: contig of 14214 bp in length
 * 157617 157660: gap of unknown length
 * 157661 172704: contig of 15044 bp in length.

FEATURES
 source
 Location/Qualifiers
 1..172704
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="Taxon:4530"
 /chromosome="10"

BASE COUNT 49372 a 35151 c 36535 g 51240 t 406 others
 ORIGIN
 /clone="OSJNBa0073L01"

Query Match 68.0%; Score 17; DB 2; Length 172704;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaattatc 24
 |||
 Db 95655 CAAGTCACAGAAATTAT 95639

RESULT 25
 AP001390/c
 LOCUS
 DEFINITION
 AP001390.2
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AP001390 174815 bp DNA HTG 30-MAY-2000
 Homo sapiens chromosome 18 clone RP11-635B11 map 18q23, WORKING
 DRAFT SEQUENCE, 21 unordered pieces.
 AP001390.2 GI:8117301
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 Homo sapiens DNA, clone:RP11-635B11.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 AUTHORS
 TITLE

1 (bases 1 to 174815)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Homo sapiens 174,815 genomic DNA of 18q23
 Published only in Database (2000) In press
 2 (bases 1 to 174815)

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Direct Submission
 Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail:hattori@psc.riken.go.jp)
 URL: http://hpg.qsc.riken.go.jp/, Tel: 81-42-778-9923,
 Fax: 81-42-778-9924
 On May 30, 2000 this sequence version replaced gi:7259325.

Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: http://hpg.qsc.riken.go.jp/
 Contact: hattori@qsc.riken.go.jp
 Project Information
 Center project name: HumDrafl18
 Center clone name: RP11-635B11

Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 159194 bases at least Q40
 Consensus quality: 157087 bases at least Q30
 Consensus quality: 170693 bases at least Q20
 Insert size: 172815; sum-of-contigs
 Quality coverage: 4.94x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
 21 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved.

1 58823 contig of 58823 bp in length
 * 58924 73243: contig of 14320 bp in length
 * 73244 83572: contig of 10229 bp in length
 * 83573 94547: contig of 10875 bp in length
 * 94548 104597: contig of 9950 bp in length
 * 104598 114347: contig of 9650 bp in length

114448 122080 contig of 7633 bp in length
 122181 128465 contig of 6285 bp in length
 128566 135791 contig of 7226 bp in length
 135892 142934 contig of 7043 bp in length
 143035 147999 contig of 4965 bp in length
 148100 151734 contig of 3635 bp in length
 151835 155576 contig of 3742 bp in length
 155677 159193 contig of 3517 bp in length
 159294 161359 contig of 2056 bp in length
 161460 163615 contig of 2056 bp in length
 163616 166614 contig of 2999 bp in length
 166715 169056 contig of 2342 bp in length
 169157 171527 contig of 2371 bp in length
 171628 173248 contig of 1621 bp in length
 173349 174815 contig of 1467 bp in length

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 58823: contig of 58823 bp in length
 58824 58923: gap of 100 bp
 58924 73243: contig of 14320 bp in length
 73244 73344: gap of 100 bp
 73344 83572: contig of 10229 bp in length
 83573 83672: gap of 100 bp
 83673 94547: contig of 10875 bp in length
 94548 94647: gap of 100 bp
 94648 104597: contig of 9950 bp in length
 104598 104697: gap of 100 bp
 104698 114347: contig of 9650 bp in length
 114348 114447: gap of 100 bp
 114448 122080: contig of 7633 bp in length
 122081 122180: gap of 100 bp
 122181 128465: contig of 6285 bp in length
 128466 128565: gap of 100 bp
 128566 135791: contig of 7226 bp in length
 135792 135891: gap of 100 bp
 135892 142934: contig of 7043 bp in length
 142935 143034: gap of 100 bp
 143035 147999: contig of 4965 bp in length
 148000 148099: gap of 100 bp
 148100 151734: contig of 3635 bp in length
 151735 151834: gap of 100 bp
 151835 155576: contig of 3742 bp in length
 155577 155676: gap of 100 bp
 155677 159193: contig of 3517 bp in length
 159194 159293: gap of 100 bp
 159294 161359: contig of 2056 bp in length
 161360 161459: gap of 100 bp
 161460 163615: contig of 2056 bp in length
 163616 166614: contig of 2999 bp in length
 166615 169056: contig of 2342 bp in length
 169057 169156: gap of 100 bp
 169157 171527: contig of 2371 bp in length
 171528 171627: gap of 100 bp
 171628 173248: contig of 1621 bp in length
 173249 173348: gap of 100 bp
 173349 174815: contig of 1467 bp in length.

FEATURES

source
 1. 174815
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q23"
 /clone="RP11-635B11"
 misc_feature
 1..58823

/note="assembly_fragment"
 58924..73243
 /note="assembly_fragment"
 73344..83572
 /note="assembly_fragment"
 83673..94547
 /note="assembly_fragment"
 94648..104597
 /note="assembly_fragment"
 104698..114347
 /note="assembly_fragment"
 114448..122080
 /note="assembly_fragment"
 122181..128465
 /note="assembly_fragment"
 128566..135791
 /note="assembly_fragment"
 135892..142934
 /note="assembly_fragment"
 143035..147999
 /note="assembly_fragment"
 148100..151734
 /note="assembly_fragment"
 151835..155576
 /note="assembly_fragment"
 155677..159193
 /note="assembly_fragment"
 159294..161359
 /note="assembly_fragment"
 161460..163615
 /note="assembly_fragment clone_end:T7 vector_side:right"
 163616..166614
 /note="assembly_fragment clone_end:SP6 vector_side:left"
 166715..169056
 /note="assembly_fragment"
 169157..171527
 /note="assembly_fragment"
 171628..173248
 /note="assembly_fragment"
 173349..174815
 /note="assembly_fragment"
 BASE COUNT 52310 a 30923 c 32183 g 57399 t 2000 others
 ORIGIN

Query Match 68.0%; Score 17; DB 2; Length 174815;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagattat 23
 DB 44384 GCAGTCACAGATTAT 44368

RESULT 26
 AC091589/c DNA HTG 26-JUL-2001
 LOCUS Homo sapiens chromosome 18 clone RP11-635B11 map 18, *** SEQUENCING
 DEFINITION IN PROGRESS *** 2 ordered pieces.
 ACCESSION AC091589
 VERSION AC091589.6 GI:15022066
 KEYWORDS HTG; HTGS-PHASE2; HTGS-FULLTOP; HTGS-ACTIVEFIN.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 178056)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 18, clone RP11-635B11
 JOURNAL Unpublished
 2 (bases 1 to 178056)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Batra,N., Bastien,V., Boguslavsky,L., Bouckhagaller,B., Brown,A.,


```

repeat_region complement(4098, .4386)
/rpl_family="LINE2"
repeat_region 4723, .4784
/rpl_family="MIR"
repeat_region 5630, .5727
/rpl_family="(CA)n"
repeat_region 7337, .7584
/rpl_family="LINE2"
repeat_region 7678, .7709
/rpl_family="(CA)n"
repeat_region 7710, .7863
/rpl_family="LINE2"
repeat_region complement(7981, .8151)
/rpl_family="LINE2"
repeat_region complement(8177, .8274)
/rpl_family="LINE2"
repeat_region complement(10118, .10192)
/rpl_family="LINE2"
repeat_region complement(10400, .10443)
/rpl_family="AT_rich"
repeat_region complement(10668, .10695)
/rpl_family="AT_rich"
repeat_region 11811, .11896
/rpl_family="LINE2"
repeat_region 12526, .12605
/rpl_family="LINE2"
repeat_region 14892, .14932
/rpl_family="MLT1A2"
repeat_region 15361, .15661
/rpl_family="AluSp"
repeat_region 16706, .16765
/rpl_family="AT_rich"
repeat_region 17193, .17764
/rpl_family="L1PA7"
repeat_region 17811, .18912
/rpl_family="L1PA7"
repeat_region complement(18968, .19018)
/rpl_family="L1"
repeat_region 19034, .19085
/rpl_family="L1"
repeat_region complement(19084, .19727)
/rpl_family="L1"
repeat_region 19737, .19971
/rpl_family="L1"
repeat_region 19971, .20406
/rpl_family="L1ME3A"
repeat_region 20452, .20500
/rpl_family="AT_rich"
repeat_region 20899, .21008
/rpl_family="(CAAAA)n"
repeat_region 21061, .21098
/rpl_family="(TAA)n"
repeat_region complement(21283, .21343)
/rpl_family="AT_rich"
repeat_region complement(21527, .22106)
/rpl_family="MLT1F"
repeat_region complement(22286, .22326)
/rpl_family="AT_rich"
repeat_region 23171, .23316
/rpl_family="MER44A"
repeat_region 23311, .23626
/rpl_family="MER44C"
repeat_region complement(23681, .23757)
/rpl_family="AT_rich"
repeat_region 24613, .24924
/rpl_family="AluSx"
repeat_region complement(24927, .25278)
/rpl_family="MSTA"
repeat_region 25554, .25668
/rpl_family="MIR"
repeat_region complement(27230, .27359)
/rpl_family="MER45"
repeat_region complement(28025, .28081)

```

```

repeat_region /rpl_family="MER45"
28227, .28312
repeat_region /rpl_family="MIR"
complement(28521, .28607)
repeat_region /rpl_family="MTR"
complement(29882, .29953)
repeat_region /rpl_family="AT_rich"
29960, .30154
repeat_region /rpl_family="MIR"
complement(30336, .30383)
repeat_region /rpl_family="AT_rich"
complement(31519, .31757)
repeat_region /rpl_family="MIR"
32198, .32905
repeat_region /rpl_family="L1MA9"
33664, .33841
repeat_region /rpl_family="MTR"
complement(34528, .34559)
repeat_region /rpl_family="AT_rich"
complement(34576, .34659)
repeat_region /rpl_family="MIR"
complement(34735, .35097)
repeat_region /rpl_family="MSTD"
complement(35249, .35546)
repeat_region /rpl_family="AluY"
complement(36693, .36767)
repeat_region /rpl_family="(CA)n"
complement(37299, .37786)
repeat_region /rpl_family="MLT1D"
38010, .38096
STS /note="L47526, HUMSNX2596, Chr. X"
/Db_xref="dbSTS: 17731"
38232, .38255
repeat_region /rpl_family="AT_rich"
39138, .39218
repeat_region /rpl_family="MIR"
complement(39901, .40077)
repeat_region /rpl_family="L1MC3"
complement(40079, .40208)
repeat_region /rpl_family="AluSg/x"
complement(40209, .40353)
repeat_region /rpl_family="L1MC3"
complement(40354, .40646)
repeat_region /rpl_family="MER47B"
complement(40647, .40798)
repeat_region /rpl_family="L1MC3"
complement(40793, .42299)
repeat_region /rpl_family="L1MC3"
complement(42297, .42391)
repeat_region /rpl_family="L1MD3"
complement(42406, .45376)

```

Query Match 68.0%; Score 17; DB 9; Length 228516;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaattattt 25
|||||

Db 55846 AAGTCACAGATTATT 55830

RESULT 28
LOCUS AP001690 340000 bp DNA PRI 30-MAY-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 34/105.
ACCESSION AP001690 AL163235 BA000005
VERSION AP001690.1 GI:7768701
KEYWORDS
SOURCE
ORGANISM Homo sapiens DNA.
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (sites)

Tag Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl₂: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

Prepared with primer pairs derived from D13644 -- Unigene.

FEATURES
source location/Qualifiers
1..1829
/organism="Homo sapiens"
STS
primer_bind 891..1001
primer_bind 891..912
primer_bind complement(977..1001)
BASE COUNT 539 a 321 c 320 g 590 t 59 others
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 1829;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacagaattat 23
|||||
Db 1689 CAAGTCACAGAATTAT 1674

RESULT 31
LOCUS D13644 4602 bp mRNA PRI 14-JUN-2001
DEFINITION Homo sapiens mRNA for KIAA0019 protein, complete cds.
ACCESSION D13644
VERSION D13644.1 GI:1531551
KEYWORDS KIAA0019.
SOURCE Homo sapiens male myoblast cell_line:KG-1 cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (sites)
TITLE Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
Prediction of the coding sequences of unidentified human genes. I.
The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
analysis of randomly sampled cDNA clones from human immature
myeloid cell line KG-1
JOURNAL DNA Res. 1 (1), 27-35 (1994)
MEDLINE 96051387
REFERENCE
AUTHORS 2 (sites)
TITLE Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
Prediction of the coding sequences of unidentified human genes. I.
The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
analysis of randomly sampled cDNA clones from human immature
myeloid cell line KG-1 (supplement)
JOURNAL DNA Res. 1 (1), 47-56 (1994)
MEDLINE 96051389

REFERENCE
AUTHORS 3 (bases 1 to 4602)
TITLE Matsukawa,B., Wong,W.T., Seki,N., Nagase,T., Nomura,N.,
Robbins,K.C. and Di Fiore,P.P.
RN-tre identifies a family of tre-related proteins displaying a
novel potential protein binding domain
onco gene 12 (12), 2563-2571 (1996)
JOURNAL Oncogene 12 (12), 2563-2571 (1996)
MEDLINE 96293402
REFERENCE
AUTHORS 4 (bases 1 to 4602)
TITLE Nomura,N., Miyajima,N., Kawarabayasi,Y. and Tabata,S.
Prediction of new human genes by entire sequencing of randomly
sampled cDNA clones
Unpublished
JOURNAL 5 (bases 1 to 4602)
REFERENCE
AUTHORS Nomura,N.
TITLE Direct Submission

JOURNAL

Submitted (11-NOV-1992) Nobuo Nomura, Kazusa DNA Research
Institute, Gene Structure 1; 1532-3 Yana, Kisarazu, Chiba 292,
Japan (E-mail:cdna@kazusa.or.jp, URL: http://www.kazusa.or.jp,
Tel:0438-52-3930, Fax:0438-52-3931)
On Sep 10, 1996 this sequence version replaced g1:286016.

COMMENT
FEATURES

source location/Qualifiers
1..4602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p13"
/sex="male"
/cell_line="KG-1"
/cell_type="myoblast"
5'UTR
gene
CDS
280..2766
/gene="KIAA0019"
280..2766
/note="protein related N-terminus of tre oncogene"
/codon_start=1
/product="KIAA0019 protein"
/protein_id="BAA02807.1"
/db_xref="GI:2104571"
/translation="MNSDQVLAQLAQRRAETVAKYDRGCAETLEPEDADLVYKY
TDRGFELHEELPDHNAVERQKLEIEETTKMLKMGWEKYNTEKFRRIYKIP
LDRGEVALLLEITPKMEETEDLSKLNHRAGSGPDIOIDVDNTEFRDIHMFRO
RYGVKQSLFHYLAAYSYIYNTGVGCGMSQITALLMNMEDAFALVLFSGPKH
AMHGRFVQGFRLRFOEHHEKILKPLSKLKHSDSEITSPYTKMFPPOCDLPT
PPTLNRITWDYIFEGENVLTAMSYTTLKLRKHLKMLSMELVFEOTLAKOFFE
DDFVLIQLOISTMLAKRALDLPKQKDEDEYPPKPLQGLPELQSGVHLSNQSV
GRPSPLASGRREGAPRRHREHSPHPOSGRTGPERAOPRRKSYEESKLIKDEADPO
RKLPSPDSSROYHNAANONSNTSNIRKEFYVKKMKPSVSATERTAKYMEGK
RAAHALATVTPGPAVGRVSNVPRKMLADADGKRGSTASOYDMVPEPDLDSGAYE
EALERAYSPPHALYPPSRPKNAEPSSPSKVSFKTKFQYOPSHARVPSOLGGEAR
GLAHPSTSNPPVYHGNSEKPPRTANSSFSASQSPGTOQLNPSRRPHSTLSVSAPE
KSYSRPSPLVPSRIEVLVDTGAGGYSNGSGSPKNGKLLIPVDYLPDNRVSEVS
YTRPEPTQOSWTRDASRGNLPKYSAGFLAPQDGLPDAVSVSPVRKASPAEDAS
PSGYPSGPPPAHYVRNDRGLSIQSEVLL"

BASE COUNT 1340 a 1041 c 999 g 1222 t
ORIGIN 3'UTR

Query Match 64.0%; Score 16; DB 9; Length 4602;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacagaattat 23
|||||
Db 4462 CAAGTCACAGAATTAT 4447

RESULT 32
LOCUS LLA012388
DEFINITION LLA012388 5469 bp DNA BCT 29-OCT-1999
Lactococcus lactis dar gene and genes encoding five hypothetical
proteins.
ACCESSION AJ012388
VERSION AJ012388.1 GI:6165401
KEYWORDS acetoin(diacetyl)reductase; dar gene; ORF1; ORF3; ORF4; ORF5; ORF6;
putative.
SOURCE Lactococcus lactis.
ORGANISM Lactococcus lactis
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Lactococcus.
REFERENCE 1 (bases 1 to 5469)
AUTHORS Aungpraphapornchai,P., Griffin,H.G. and Gasson,M.J.
TITLE Cloning, DNA sequence analysis, and deletion of a gene encoding
diacetyl-acetoin reductase from Lactococcus lactis
JOURNAL DNA Seq. 10 (3), 163-172 (1999)
MEDLINE 20112344
REFERENCE 2 (bases 1 to 5469)

AUTHORS
TITLE
JOURNAL

Aungmyaphapornchai, P.
Direct Submission
Submitted (28-Oct-1998) Aungmyaphapornchai P., Genetics and
Microbiology Institute of Food Research, Colney Lane, Norwich,
Norfolk, NR4 7UA, UK

FEATURES
Source

Location/Qualifiers
1..5469

/organism="Lactococcus lactis"
/sub_species="cremoris"

/strain="MG1363"

/db_xref="taxon:1358"

77..80

/note="ORF1"

87..950

/note="ORF1"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="CAB59825.1"

/db_xref="GI:6165402"

/translation="MNPKNRIILAVAVLILVALVAFSLNHOGVAKASGEKTVAVG
IMSGDQDOEYKMSVANTAKEXYDLKLFVYFSDYNOPEALISCDIINAROSYVY
KTMKAKRSDIVAGNTYITPMHYSKESIKSLDKCSTVALPNDASNSRALFVYQ
SAGILKLTSDSSKLVLPDITENPHOLKREVDASOTPAALSVASLVNVAITVA
SLPKSEVFMPELKRISAOYINFLATTSKERNKVKYKAVAKASATEKAIEQYPD
GGLPAMDLLK"

1005..1009

/gene="dar"

1005..1789

/gene="dar"

1019..1789

/gene="dar"

/codon_start=1

/transl_table=1

/product="diacetyl (acetoin) reductase"

/protein_id="CAB59826.1"

/db_xref="GI:6165403"

/translation="MCKLAIVTSGGICIGTYRIKLAKGFAVAADLNPQAKNNAK
TLIDGFEAKAYVLDANRSVAFOLIEOAVSDGELAVFNAGVAFIDSIIDSDER
VERLEFDVNIKGTWGIQAAATOPKQOHGRIINAAISVSGSALQSAVASAKFAIK
GLTQSAKELAKDITVNAVNPGLVIRTAMDIEDKRTAQIKGTISIEQOONCLNELYL
GREGPEDEVAEVAVMFASDAEYITGOSILVDCMRPN"

1794..1798

/note="ORF3"

1805..2665

/note="ORF3"

/codon_start=1

/transl_table=1

/product="hypothetical proteia"

/protein_id="CAB59827.1"

/db_xref="GI:6165404"

/translation="MNYKNRILITIIILVFIIVGIGFAFSHGSKSVSSKIVKIG
LMPGKQEDYIMKOVONAKDQGITLKVNPFDGDEPKNALNHNHVDLNAQOHAVL
KSNKANGNIVISIGDITITPILHLYSTKVKKDEIDPSTIALPNDITESHALVYL
NAGLIKIDTSRGVATVKDIRENPKSLIKEIDAPSPALDSVAAVINNYFAISAK
NSDKESIOEPLNEDSACWINPIAANQSDNNKVVYKEVAVKAEQKNIAIDIIKEPDDG
GELPAMNLKL"

2776..2781

/note="ORF4"

2789..3895

/note="ORF4"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="CAB59828.1"

/db_xref="GI:6165405"

/translation="MTAILLNNLSVOFHQKRLVAVKNAATLHIEKGDYVIGVYSG
AGKSTLVRITNLQKPTGEOIVINKEIPDENPVFTGAKLREFROKIGMLEFHNIL
LSKTYENNVAFALQSDIEDKNGKRYLTKKKNKXVTELLIKVIDIADLSKRYAOL
SGGOKORVALIARLANDPETLISDPECTSALDEKPTNLLDLKLSHEKIKTIVTILTH
EMQVYEIAKNAVAMONGETIEONSILIDIPADPKFALIKQFETTSVNRPTASLSKT
ELLAQIADDEELIHIDYSGSELEDVPSVDTKKFVPTNIPVGNVELLQGGPQSLVL
TIKGSSEHRAAKAVYFVERHLKFEVYLGKIERITVD"

RBS
CDS

RBS

3878..3884

/note="ORF5"

3895..4590

/note="ORF5"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="CAB59829.1"

/db_xref="GI:6165406"

/translation="MAEWFAHFPNVVYLGMTGETGMWTAIVOLYMTFISALIGGL
GLIFGIGVAVTAEDGITPNRPLMLDKITVSIGRAFPTILAAIAPLTKLVGTQIG
VTALVPLALGAPFYAROVQASLESVDHGKVAQVGADEFLDVFVYREFEASL
IIVSTVTLISLIGTAMAGATGAGIGLMTAISYGNBRANDVMTFATLILIFLLVQ
LVGDFLARVSHR"

4621..4627

/note="ORF6"

4633..5181

/note="ORF6"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="CAB59830.1"

/db_xref="GI:6165407"

/translation="MKNSVKIVATGIGALFVILGMLINIPPIPTSIOLQAVL
ALFSAIFGFLAGFLIGTIGHALKDSFLGAPMMTWLSSGLMGFLGCVRESLTGG
IFGNKEITRFNIVQFLANVWVWGILAPIGIDILVYSEPAKVKFTGGVAGLVNALTIV
ACTLLKLYAATRTKSGFLDKE"

5210..5235

/note="putative"

terminator

BASE COUNT

1796 a

858 c

1134 g

1681 t

ORIGIN

Query Match

Best Local Similarity

100.0%; Score 16;

DB 1;

Length 5469;

Matches 16;

Conservative 0;

Mismatches 0;

Indels 0;

Gaps 0;

QY

9 aagtcacagaatatt 24

|||||

Db

3204 AAGTCACAGAATATT 3219

RESULT 33

AL161800/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human

Human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 43741)

Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequests@sanger.ac.uk

On Jun 4, 2000 this sequence version replaced gi:8217673.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Emi., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP11-318C17 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP5-974N19 is at 43642 in this sequence. The true right end of clone RP5-858M22 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats: all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-318C17 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

FEATURES

source

1. 43741
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP11-318C17"
/clone_1lb="RP11-11.2"
3. 293
/note="AluX repeat: matches 1. 291 of consensus"
390. 683
/note="AluSg repeat: matches 1. 286 of consensus"
788. 3327
/note="L1PA16 repeat: matches 3587. 6157 of consensus"
3411. 3820
/note="match: GSS: Em:AQ371040"
4157. 4477
/note="L2 repeat: matches 2348. 2710 of consensus"
/note="match: GSS: Em:AQ236127"
/note="match: GSS: Em:AQ236127"
complement(6607. 6950)
/note="match: GSS: Em:AQ110431"
6676. 6820
/note="MIR repeat: matches 79. 229 of consensus"
7231. 7405
/note="MIR repeat: matches 42. 206 of consensus"
7416. 7551
/note="L2 repeat: matches 2584. 2710 of consensus"
9521. 9605
/note="MLT1G repeat: matches 433. 523 of consensus"
/note="join(9542. 9883,13385. >135933))"
/gene="BA318C17.1"
/note="continued from d1974N19.2 in Em:AL138808 match: ESTs: Em:AL542749 Em:AL565992"
/evidence="not_experimental"
/product="BA318C17.1 (novel transcript)"
/complement(9542)
/gene="BA318C17.1"
complement(join(9542. 9883,13385. 135933))
/gene="BA318C17.1"
complement(9565. 9570)
/gene="BA318C17.1"
9805. 9884
/note="MLT1G repeat: matches 59. 137 of consensus"
9885. 10330
/note="MR4B repeat: matches 1. 437 of consensus"
10325. 10397
/note="MR4C repeat: matches 393. 465 of consensus"
10682. 11085
/note="L2 repeat: matches 1981. 2406 of consensus"
11167. 11657
/note="L2 repeat: matches 2263. 2750 of consensus"
11872. 12002
/note="L2 repeat: matches 2621. 2750 of consensus"
12681. 12976

repeat_region
14480. 15774
/note="L2 repeat: matches 1424. 2749 of consensus"
15775. 16067
/note="AluX repeat: matches 1. 310 of consensus"
16068. 16216
/note="L2 repeat: matches 1285. 1424 of consensus"
16713. 16865
/note="L2 repeat: matches 2545. 2699 of consensus"
17484. 17764
/note="AluX repeat: matches 6. 296 of consensus"
17813. 18013
/note="L1M2 repeat: matches 5966. 6164 of consensus"
18014. 18267
/note="AluSc repeat: matches 1. 252 of consensus"
18268. 18773
/note="L1M2 repeat: matches 5450. 5966 of consensus"
19067. 19374
/note="AluSg1 repeat: matches 1. 304 of consensus"
19513. 19811
/note="AluY8 repeat: matches 1. 305 of consensus"
21304. 21613
/note="AluSg repeat: matches 1. 308 of consensus"
22998. 23343
/note="MLT1A1 repeat: matches 1. 353 of consensus"
23736. 23865
/note="AluSg repeat: matches 1. 130 of consensus"
24570. 25025
/note="MLT1C repeat: matches 1. 466 of consensus"
25235. 26106
/note="L1M4 repeat: matches 3073. 4010 of consensus"
complement(26577. 27248)
/note="match: GSS: Em:AQ36701"
28315. 28629
/note="AluSg repeat: matches 1. 313 of consensus"
28933. 29096
/note="MLT1 repeat: matches 245. 409 of consensus"
28993. 30110
/note="MIR repeat: matches 25. 245 of consensus"
complement(30032. 30494)
/note="match: GSS: Em:AQ23457"
31501. 31807
/note="AluX repeat: matches 1. 307 of consensus"
32165. 32267
/note="L2 repeat: matches 2554. 2646 of consensus"
32268. 32752
/note="MR6B repeat: matches 1. 486 of consensus"
32753. 33089
/note="L2 repeat: matches 2213. 2554 of consensus"
33351. 33651
/note="AluX repeat: matches 1. 304 of consensus"
33662. 33966
/note="AluY repeat: matches 1. 295 of consensus"
34730. 34818
/note="MIR repeat: matches 43. 145 of consensus"
complement(34862. 35287)
/note="match: GSS: Em:AQ404224"
complement(35239. 35778)
/note="match: GSS: Em:AQ634531"
complement(35374. 35798)
/note="match: GSS: Em:AQ565906"
36173. 36533
/note="match: GSS: Em:AQ231637"
36392. 36553
/note="MR5A repeat: matches 4. 175 of consensus"
36668. 36961
/note="AluSg repeat: matches 1. 297 of consensus"
37033. 37258
/note="MIR repeat: matches 15. 252 of consensus"
38080. 38507
/note="match: GSS: Em:B36377"
38087. 38612
/note="match: GSS: Em:AQ680468"

```

repeat_region 38349..38418
/note="33 copies 2 mer tg 748 conserved"
repeat_region 38694..38979
/note="AluId repeat: matches 1..286 of consensus"
misc_feature 40625..41129
/note="match: GSS: Em:A0680145"
repeat_region 41704..41819
/note="MIR repeat: matches 57..177 of consensus"
repeat_region 41842..42034
/note="L1MC4 repeat: matches 7779..7969 of consensus"
repeat_region 42037..42226
/note="MLT1J repeat: matches 45..247 of consensus"
misc_feature complement(42848..43207)
/note="match: GSS: Em:A0385536"
repeat_region 42966..43238
/note="L2 repeat: matches 1138..1493 of consensus"
repeat_region 43259..43555
/note="AluId repeat: matches 10..303 of consensus"
repeat_region 43556..43741
/note="L2 repeat: matches 970..1138 of consensus"
BASE COUNT 12960 a 8117 c 8797 g 13867 t
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 9; Length 43741;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 9 aagtcacgaattatt 24
|||||
Db 10623 AAGTCACGAATTATT 10608

```

```

RESULT 34
AC074073 68894 bp DNA HTG 13-JUL-2000
LOCUS Homo sapiens chromosome 3 clone RP11-79116 map 3, LOW-PASS SEQUENCE
DEFINITION AC074073
ACCESSION AC074073
VERSION AC074073.1 GI:9170900
KEYWORDS HTG; HTGS-PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 68894)
AUTHORS Birren,B., Linton,J., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 3, clone RP11-79116
TITLE Unpublished
AUTHORS 2 (bases 1 to 68894)
Birren,B., Linton,J., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barrera,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
Campoliano,A., Casale,A., Choepel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hacos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPherson,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Miñana,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisanu,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talmas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,
Vassiliev,H., Vleij,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

```

```

TITLE Direct Submission
JOURNAL Submitted (13-JUL-2000) Whitehead Institute/MIT Center for Genome

```

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information -----
 Center project name: L7127
 Center clone name: 791_L16

* NOTE: This record contains 83 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

```

1 747: contig of 747 bp in length
1 748 847: gap of 100 bp
1 848 1579: contig of 732 bp in length
1 1580 1679: gap of 100 bp
1 1680 2425: contig of 746 bp in length
1 2426 2525: gap of 100 bp
1 2526 3254: contig of 729 bp in length
1 3255 3354: gap of 100 bp
1 3355 4075: contig of 721 bp in length
1 4076 4175: gap of 100 bp
1 4176 4892: contig of 717 bp in length
1 4893 4992: gap of 100 bp
1 4993 5717: contig of 725 bp in length
1 5718 5817: gap of 100 bp
1 5818 6558: contig of 741 bp in length
1 6559 6658: gap of 100 bp
1 6659 7371: contig of 713 bp in length
1 7372 7471: gap of 100 bp
1 7472 8209: contig of 738 bp in length
1 8210 8309: gap of 100 bp
1 8310 9046: contig of 737 bp in length
1 9047 9146: gap of 100 bp
1 9147 9888: contig of 742 bp in length
1 9889 9988: gap of 100 bp
1 9989 10723: contig of 735 bp in length
1 10724 10823: gap of 100 bp
1 10824 11553: contig of 730 bp in length
1 11554 11653: gap of 100 bp
1 11654 12391: contig of 738 bp in length
1 12392 12491: gap of 100 bp
1 12492 13214: contig of 723 bp in length
1 13215 13314: gap of 100 bp
1 13315 14034: contig of 720 bp in length
1 14035 14134: gap of 100 bp
1 14135 14864: contig of 730 bp in length
1 14865 14964: gap of 100 bp
1 14965 15700: contig of 736 bp in length
1 15701 15800: gap of 100 bp
1 15801 16535: contig of 735 bp in length
1 16536 16635: gap of 100 bp
1 16636 17357: contig of 722 bp in length
1 17358 17457: gap of 100 bp
1 17458 18192: contig of 735 bp in length
1 18193 18292: gap of 100 bp
1 18293 19030: contig of 738 bp in length
1 19031 19130: gap of 100 bp
1 19131 19857: contig of 727 bp in length
1 19858 19957: gap of 100 bp
1 19958 20677: contig of 720 bp in length

```

```

* 20678 20777: gap of 100 bp
* 20778 21514: contig of 737 bp in length
* 21515 21614: gap of 100 bp
* 21615 22336: contig of 722 bp in length
* 22337 22436: gap of 100 bp
* 22437 23190: contig of 754 bp in length
* 23191 23290: gap of 100 bp
* 23291 24024: contig of 734 bp in length
* 24025 24124: gap of 100 bp
* 24125 24859: contig of 735 bp in length
* 24860 24959: gap of 100 bp
* 24960 25700: contig of 741 bp in length
* 25701 25800: gap of 100 bp
* 25801 26544: contig of 744 bp in length
* 26545 26644: gap of 100 bp
* 26645 27367: contig of 723 bp in length
* 27368 27467: gap of 100 bp
* 27468 28196: contig of 729 bp in length
* 28197 28296: gap of 100 bp
* 28297 29028: contig of 732 bp in length
* 29029 29128: gap of 100 bp
* 29129 29843: contig of 715 bp in length
* 29844 29943: gap of 100 bp
* 29944 30633: contig of 690 bp in length
* 30634 30733: gap of 100 bp
* 30734 31472: contig of 739 bp in length
* 31473 31572: gap of 100 bp
* 31573 32318: contig of 747 bp in length
* 32320 32419: gap of 100 bp
* 32420 33151: contig of 732 bp in length
* 33152 33251: gap of 100 bp
* 33252 33977: contig of 726 bp in length
* 33978 34077: gap of 100 bp
* 34078 34818: contig of 741 bp in length
* 34819 34918: gap of 100 bp
* 34919 35632: contig of 714 bp in length
* 35633 35732: gap of 100 bp
* 35733 36474: contig of 742 bp in length
* 36475 36574: gap of 100 bp
* 36575 37300: contig of 726 bp in length
* 37301 37400: gap of 100 bp
* 37401 38137: contig of 737 bp in length
* 38138 38237: gap of 100 bp
* 38238 38960: contig of 723 bp in length
* 38961 39060: gap of 100 bp
* 39061 39790: contig of 729 bp in length
* 39790 39889: gap of 100 bp
* 39890 40617: contig of 728 bp in length
* 40618 40717: gap of 100 bp
* 40718 41442: contig of 725 bp in length
* 41443 41542: gap of 100 bp
* 41543 42278: contig of 736 bp in length
* 42279 42378: gap of 100 bp
* 42379 43117: contig of 739 bp in length
* 43118 43217: gap of 100 bp
* 43218 43961: contig of 744 bp in length
* 43962 44061: gap of 100 bp
* 44062 44786: contig of 725 bp in length
* 44787 44886: gap of 100 bp
* 44887 45614: contig of 728 bp in length
* 45615 45714: gap of 100 bp
* 45715 46458: contig of 744 bp in length
* 46459 46558: gap of 100 bp
* 46559 47292: contig of 734 bp in length
* 47293 47392: gap of 100 bp
* 47393 48107: contig of 715 bp in length
* 48108 48207: gap of 100 bp
* 48208 48935: contig of 728 bp in length
* 48936 49035: gap of 100 bp
* 49036 49765: contig of 730 bp in length
* 49766 49865: gap of 100 bp
* 49866 50604: contig of 739 bp in length
* 50605 50704: gap of 100 bp

```

```

Query Match      64.0%; Score 16; DB 2; Length 68894;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 agtcacagattatc 25
Db 58360 AGTCACAGATTATT 58345
|||||

```

```

RESULT 35
AL590387/c      76075 bp      DNA      PRI      06-JUN-2001
LOCUS
DEFINITION
Human DNA sequence from clone RP3-356B20 on chromosome 6, complete
sequence.
ACCESSION      AL590387
VERSION
KEYWORDS
SOURCE
ORGANISM
human.

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

COMMENT
Direct Submission
Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced GI:14141468.

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EM: EMBL, SW: SWISSPROT, Tr: TrEMBL, Wp: WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

RP3-356B20 is from the library RP3-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2
 IMPORTANT: This sequence is not the entire insert of clone RP3-356B20 It may be shorter because we sequence overlapping


```

repeat_region 41564..42004
                /note="L2 repeat: matches 2256..2750 of consensus"
repeat_region 42012..42110
                /note="L2 repeat: matches 2587..2693 of consensus"
repeat_region 42443..42526
                /note="42 copies 2 mer at 63% conserved"
repeat_region 42537..42787
                /note="L1MD2 repeat: matches 6108..6341 of consensus"
repeat_region 42853..43106

Query Match      64.0%; Score 16; DB 9; Length 76075;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattattc 25
    |||||
Db 17122 ACTCAGAGATTATTT 17107

RESULT 36
AC012679/c 105937 bp DNA PLN 19-JAN-2001
LOCUS Arabidopsis thaliana chromosome 1 BAC F25P22 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC012679
VERSION AC012679.5 GI:12324196
KEYWORDS HMG;
SOURCE Thale cress.
ORGANISM Arabidopsis thaliana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
           1 (bases 1 to 105937)
           Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.J.,
           Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
           Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
           Arabidopsis thaliana chromosome 1 BAC F25P22 genomic sequence
           Unpublished
           2 (bases 1 to 105937)
           Lin,X. and Kaul,S.
           Direct Submission
           Submitted (03-NOV-1999) The Institute for Genomic Research, 9712
           Medical Center Dr, Rockville, MD 20850, USA, xlin@ligr.org
           3 (bases 1 to 105937)
           Town,C.D. and Kaul,S.
           Direct Submission
           Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
           Medical Center Dr, Rockville, MD 20850, USA, cdtown@ligr.org
           On Jan 19, 2001 this sequence version replaced gi:12280804.
           Address all correspondence to:at@ligr.org

COMMENT
BAC clone F25P22 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from Sp6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://cgr-081.mit.edu/GENSCAN.html), GenemarkHMM (Mark Borodovsky,
http://genomebiology.gatech.edu/genemark/), Glimmer4 (a variant
of Glimmer3, see Mihalea Pertea,
http://www.tigr.org/softlab/glimmer4/glimmer4.html, and
GeneSplicer (Mihalea Pertea and Steven Salzberg, contact
mperle@ligr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smit,

```

```

FEATURES
    source
        http://ftp.genome.washington.edu/RM/RepeatMasker.html).
        Location/Qualifiers
            1..105937
                /organism="Arabidopsis thaliana"
                /cultivar="Columbia"
                /db_xref="taxon:3702"
                /chromosome="1"
                /map="nga111"
                /clone="F25P22"
                /join<1..76,158..299,430..710,789..950,1031..1084,
                1171..1512)
                /gene="F25P22.1"
                /note="F25P22.1"
                /note="similar to
                S-adenosyl-methionine-sterol-C-methyltransferase
                GB:AC34951 [Nicotiana tabacum]
                join<1..76,158..299,430..710,789..950,1031..1084,
                1171..1344)
                /gene="F25P22.1"
                /codon_start=2
                /product="putative
                S-adenosyl-methionine-sterol-C-methyltransferase, 5'
                partial; 1-1344"
                /protein_id="AAC52075.1"
                /db_xref="GI:12324209"
                /translation="NSVELSLVSKCIGAYVRNKNQNOICMLQKVSNDRGFORF
                LDVQKSSGILRYERFEGFVSTGGLTTEFPDMDLKRQGVLDVCGIGGDF
                YMAENFDVVDVGLDLSVNMISFLEHAICKCSVEFEVADCTKKEIPDVTEDVTSRD
                TLIHIDOKPALFRFRFYKLPQGVKLTIDYCRSPKPSDFAIYTKRGYDLHDOAY
                GQMLRACGEVLEAEDRTDQFMKVLKRELDAAVEKEEFISDFSEDEDIIGKMSK
                LERSSGEQKMGFLIKRN"
                join<6321..6579,6690..6820,6910..7146,7259..7502,
                7588..7751)
                /gene="F25P22.2"
                /gene="F25P22.2"
                /note="contains Pfam profile: PF00657 lipase/acylhydrolase
                with GDSL-like motif"
                join(6321..6579,6690..6820,6910..7146,7259..7502,
                7588..7751)
                /gene="F25P22.2"
                /codon_start=1
                /product="putative lipase/acylhydrolase; 6321-7751"
                /protein_id="AAC52082.1"
                /db_xref="GI:12324216"
                /translation="MNCIMFRKMLLAFSISLFYVGNAGQSGNSTVSLFAFGDSIL
                DTGNNDLISVSVNFTPTGRDFTGRATGRGNGRVSDITAEGLKNLPRAYRDP
                YLANNNDLTGVCVPSGSGSLDPTTARTGSIWSDQVDFQNTITRLNGVGNQDQAN
                AVISNAVYLISAGNNDIATITFTGARLQVTLPAVNDQLSWTRDLKSLDKMGRK
                FAVMGTLPLGCLPGARALTRACELFVNQGAAMNQLSDADINLGTGTFGAFVYVDK
                YNLLGLITINPQASGFLIDVADACCPHTLIPCLASRYVEMDVAPHQKSVETIAPQ
                ILENIKAKLIA"
                join(<9376..9496,9643..9967,10550..>10898)
                /gene="F25P22.3"
                /gene="F25P22.3"
                /note="contains Pfam profile: PF00314 thaumatin family"
                join(9376..9496,9643..9967,10550..10898)
                /gene="F25P22.3"
                /codon_start=1
                /product="thaumatin-like protein; 9376-10898"
                /protein_id="AAG52086.1"
                /db_xref="GI:12324220"
                /translation="MTNKEKVALPRLKISEKSMASSSSLFLPLLLLSVQASASTV
                IFYNAKTYVWPDIOPSSGSLTAGGFFLSPKRAATLQPLPLMSRFMRGRCSDR
                SGGRKATGCGGSLFCLNGAGVPPATLEITLIGHMDYDLSLVGVMLANSIMPKV
                GTGKCTAGAGSDSLNRMCPVGLDVRSDGTQVAVACSKASAFNSPCTCTGFLGNVS
                CKPVAISKIRKVCAPKAYSAYVDDPSIATGCSKANVVFPCPHRG"
                <12692..>13183
                /gene="F25P22.4"
                /gene="F25P22.4"
                /gene="F25P22.4"

```

/note="similar to calmodulin GB:P02596 [Renilla reniformis]: contains Pfam profile: PF00036 EF hand (4 copies)"

CDS
12692..13183
/gene="F25P22.4"
/codon_start=1
/product="putative calmodulin; 12692-13183"
/db_xref="GI:12324222"
/protein_id="AAG52088.1"
/translation="MAANTLESTNKSTTPSTDMELKVFDEKFDANGDCKLSVSELGV
FKSGTSTEEELNKVDEIDIDCGFINOEFAATICSSSAVEIRAPDLDOONN
GLISSEIHKVNLRLGMCVSDCVRMIGHVDTGDDGVNVEEFQKMMSPSELVKGV
ANS"

mRNA
join(<14977..15197,15451..>15931)
/gene="F25P22.5"
14977..15931
/note="contains Pfam profile: PF00071 ras family"
join(14977..15197,15451..15931)
/gene="F25P22.5"
/codon_start=1
/product="putative ras-related GTP-binding protein;
14977-15931"

gene
/protein_id="AAG52089.1"
/db_xref="GI:12324223"
/translation="MAEDYEEDCYLEKAVLIGDSAVGKSNLRSFSDPEFEDSKP
TIGEFARVAVGDKIIKAOIMPTAGERRAITSSYRCALGALLIYITRTTFD
NIKKLPFLPDPANPEYVYLVCNKSRLROREVEDEDEKTLASEDLTELETSALIN
VNVEAPILVIGRIHEVYTORIASEKNSNGAITHINGNGVILPVGKEIVNHEVI
ATPLSSSSWCKE"

mRNA
complement(join(<16797..16967,17064..17297,17388..17469,
17546..17607,17706..17773,17866..17935,18448..18772,
19179..19223,19427..>19725))
/gene="F25P22.7"
complement(16797..19725)
/gene="F25P22.7"
complement(join(16797..16967,17064..17297,17388..17469,
17546..17607,17706..17773,17866..17935,18448..18772,
19179..19223,19427..19725))
/gene="F25P22.7"
/codon_start=1
/product="unknown protein; 19725-16797"
/protein_id="AAG52066.1"
/db_xref="GI:12324200"
/translation="MATFLATVAFVSHRFVSPQHPKOSLQSLSVTFTEPNQPTAV
VTLQEOQLDWTSPVIRFEGIFAGFTWAGLAFGVSEDMKSRDLVPEEENITGL
EKOEIILPNCIPIRYDLOVSGATPSSGYLVFPVKGAVGTQVDFVDEGGKSLA
MKMSRPSYSGKICGIEHVLRSKAGGKRRVITPPSGFEDRVEREGGLEIPPSATL
DYILEVIXQEFPIYITLAFKFDVTDAGSTNYITLAVLTLKASHEKQIVLTILV
VWGRLGIFELMKILOWGEDRDEBORGNIIVRIITFVLQAVWVTVSLPLLVNAS
DGGSLKPADVIGMTVMVGFLEAADOQKLSFKNSPENRGWCDGVKWSLRHPNY
FGEMILMWGIFVAAAPVLEGAEVIVIFGPFLLLLFLFVSGIPLLEVLPCSLPLYLDN
P"

mRNA
complement(join(<20191..20329,20451..20627,20721..20770,
20870..20938,21048..21146,21239..21317,21582..21682,
21776..21844,21957..21997,22081..23243,23382..23485,
23568..24008,24102..>24662))
/gene="F25P22.8"
complement(20191..24662)
/gene="F25P22.8"
/note="contains Pfam profile: PF00069 eukaryotic protein
kinase domain"
complement(join(20191..20329,20451..20627,20721..20770,
20870..20938,21048..21146,21239..21317,21582..21682,
21776..21844,21957..21997,22081..23243,23382..23485,
23568..24008,24102..24662))
/gene="F25P22.8"
/codon_start=1
/product="putative protein kinase; 24662-20191"
/protein_id="AAG52069.1"
/db_xref="GI:12324203"
/translation="MKVEETLKNLGDGVVLRPVDGSSISWKRMMKMKFLKLIASP
NOSDEAGSISTTKSNHKKSIDVSSSSSPRSHHSNPEIKPFGSLSNWLSVSGHRIP

SPNSRMAKNRATVDDTYVNGSEHVDGSDPAVEEENOIQLLEISARDEPRATO
IEATKOPSLGASCAPENSPADLLATRYWNLGLDYDKILDGFYDLYGLNASSABERIP

Query Match 64.0%; Score 16; DB 8; Length 105937;
Best local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 10 agtcacgaattattc 25
|||||
Db 55362 AGTCACGAATTATT 55347

RESULT 37
AC044784_1/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC044784 Accession AC044784
Fragment Name Begin End
AC044784.0 1 110000
AC044784.1 100001 210000
AC044784.2 200001 310000
AC044784.3 300001 365176
Continuation (2 of 4) of AC044784 from base 100001 (AC044784 Homo sapiens chromosome

Query Match 64.0%; Score 16; DB 2; Length 110000;
Best local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 8 caagtcacgaattat 23
|||||
Db 10285 CAACTCACGAATTAT 10270

RESULT 38
AB020878
LOCUS
DEFINITION
Homo sapiens genomic DNA of 9q32 anti-oncogene of flat epithelium
cancer, segment 10/10.
ACCESSION
AB020878
VERSION
AB020878.1 GI:4003398
KEYWORDS
Homo sapiens DNA.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (sites)
Nakamura, Y., Isomura, M. and Daigo, Y.
Homo sapiens 1,010,525bp genomic DNA of 9q32 anti-oncogene of flat
epithelium cancer region
Published Only in Database (1998) In press
2 (bases 1 to 110525)
Hirakawa, M., Yamaguchi, H., Imai, K. and Shinada, J.
Direct Submission
Submitted (30-NOV-1998) to the DDBJ/EMBL/GenBank databases. Mka
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Database Department, 5-3, Yonban-cho, Chiyoda-ku, Tokyo 102-0028,
Japan (E-mail:mika@tokyo.jst.go.jp, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
This sequence is conducted by Japanese Foundation for Cancer
Research as a JST seauencing team
Principal Investigator: Yusuke Nakamura Ph.D.
Phone:+81-3-5449-5372, Fax:+81-3-5449-5433,
yusuke@jst.go.jp
The sequence is submitted by:human Genome Sequencing in ATIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site
(http://www-atis.tokyo.jst.go.jp/HGS/top.html)
or send email to webmaster@www-atis.tokyo.jst.go.jp.
Location/Qualifiers

FEATURES

```

source
1. 110525
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="9"
   /map="9q32"
   /start="80338"
   /end="80503"
   /standard_name="D9S1126"
   /note="CHLC.GGA11B01.P33342:The location is between each
   flanking site of PCR primers."
   /db_xref="GDB:686181"
   /db_xref="GDB:686181"

STS
64.0%; Score 16; DB 9; Length 110525;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaattatt 24
DB 39879 AACCTACAGATTATT 39894

RESULT 39
AC011248 111312 bp DNA PRI 28-APR-2000
LOCUS Homo sapiens 12 BAC RP11-14P21 (Roswell Park Cancer Institute Human
DEFINITION BAC library) complete sequence.
AC011248
AC011248.8 GI:7658299
HTG:
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 111312)
REFERENCE
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodoc,B., Boucek,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guerrero,W., Harris,K., Hernandez,J., Hodgson,A., Hughes,M.,
Holloway,C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,
Jones,M., Kelly,B., Lee,E., Li,Z., Lichtarge,O., Liu,J., Liu,W.,
Lau,S., Leal,B., Lee,S., Kneitz,S., Kondrjewski,N., Kong,Y., Kovar,C.,
Logan,O., Lozano,R.J., Lu,J., Lucier,R., Marondel,I., Martin,R.,
Martinez,C., McLeod,M.P., Mel,G., Merscher,S., Miller,A.,
Montgomery,K.T., Morgan,M., Morris,S., Nash,S., Nelson,A.,
Nguyen,R., Nguyen,N., Nguyen,S., Oswald,G., Parish,B., Paxton,S.,
Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M.,
Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Shim,C., Simon,M.,
Sparks,A., Stamps,A., Sugeng,R., Taber,P., Taylor,T., Vasquez,L.,
Vinson,R., Vo,O., Wahbah,M., Watlington,S., Weinstein,G.,
Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G.,
Xiang,A.M., Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 111312)
Worley,K.C.
Direct Submission
Submitted (05-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
3 (bases 1 to 111312)
Worley,K.C.
Direct Submission
Submitted (28-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2000 this sequence version replaced g1:6969601.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
html.

QUALSTAT-REPORT-----
----- Summary Statistics -----
Contig Length: 150286
Phrap values in estimate: 149476
Average error rate (BCM-Phrap estimate): 0.000180006
Fraction of Phrap values less than 40 : 0.0291418
Number of consensus changing edits: 15
Number of N's in consensus : 1

----- Consensus changing edits -----
Position Original+Context Edited+Context
22441 aaagaaga(n)gaggtccca aaagaaga(a)gaggtccca
79231 agtaagcca(n)annttann agtaagcca(t)atttaagc
79233 taagccana(n)tttanncc taagccata(t)tttaagtc
79234 aggcacana(n)tttanncc aggcacata(t)tttaagtc
79235 caaanltta(n)lncctgtg caaanltta(a)tgctgtgta
79241 nannlttant(n)ccctgtgaa tttttat(g)ctgtgtgaa
79245 ttantncc(n)gtgaaataaa ttatgtcc(t)gtgaaataa
79279 agagctgtga(n)gtcagaatg agagctgtga(a)gtcagaatg
79440 agcctgtgca(n)catcgtgaa lctcaaaaa(a)caaaagaaag
89920 lctcaaaaa(n)caaaagaaag lctcaaaaa(a)caaaagaaag
94117 gtttccag(t)tcctgtgaca gtttccag(t)tcctgtgaca
95790 ggcagattcc(n)lctcaaaaa ggcagattcc(a)lctcaaaaa
111017 ttaagttcc(n)ltagctgaaac ttaagttcc(a)ltagctgaaac
115887 atgccata(n)caacaacaa atgccata(a)caacaacaa
129846 ggagctgaa(n)cgagtgatc ggagctgaa(g)cgagtgatc

```

```

----- Distribution of Quality < 40 Bases -----
#
bases
1000
900
800
700
600
500
400
300
200
100
0
* * * * *
5 10 15 20 25 30 35 40
Phrap Value Range

```

```

FEATURES
source
Version: 1.01 gxf.
Location/Qualifiers
1. .11312
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-14P21"
1. .133
/rpt_family="Alusg/x"
167. .193
/rpt_family="(GACA)n"
553. .795
/rpt_family="MIR"
complement(824. .901)
/rpt_family="MIR"
946. .985
/rpt_family="(CA)n"
complement(1467. .1648)
/rpt_family="MLTIG"
1876. .2176
/rpt_family="Alusq"
complement(2180. .2431)
/rpt_family="Alusx"
complement(2432. .2788)
/rpt_family="HEIB"
complement(2789. .2835)
/rpt_family="Alusx"
complement(2864. .3025)
/rpt_family="MLTIF1"
3101. .3327
/rpt_family="MIR"
complement(3383. .3455)
/rpt_family="L2"
3974. .4064
/standard_name="G07716"
/db_xref="dbSTS:13562"
5179. .5480
/rpt_family="AlusP"
complement(6278. .6347)
/rpt_family="MER4A"
6348. .6648
/rpt_family="Alusg"
complement(6649. .7184)
/rpt_family="MER4A"
8486. .8703
/rpt_family="L2"
8729. .9122
/rpt_family="L2"
9213. .9497
/rpt_family="L2"
9498. .9798
/rpt_family="Aluy"

```

```

repeat_region 9799. .9887
/rpt_family="L2"
repeat_region complement(10707. .12039)
/rpt_family="LIM4"
repeat_region 12047. .12155
/rpt_family="L2"
repeat_region complement(12165. .12729)
/rpt_family="L2"
repeat_region 12702. .12847
/rpt_family="L2"
STS 13332. .13486
/standard_name="Cd11f102"
/db_xref="GDB:446262"
repeat_region 13505. .13528

```

```

Query Match 64.0%; Score 16; DB 9; Length 111312;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 aagtcacagatatt 24
DB 86830 AACTCACAGATTATT 86845
|||||
RESULT 40
HSDJ655C5 119733 bp DNA PRI 03-JUL-2001
LOCUS Human DNA sequence from clone RP4-655C5 on chromosome 6, complete
DEFINITION sequence.
ACCESSION AL121956
VERSION AL121956.21 GI:14599397
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 119733)
REFERENCE
AUTHORS Almeida J.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 4, 2001 this sequence version replaced gi:14456158.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: EM, EMBL, SW,
SWISSPROT, Tr, TREMBL, Wp, WORMPEP. Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/RRP/Chr6
RP4-655C5 is from the library RP4-4 constructed by the group of
Pleier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP4-655C5.
FEATURES
source
Location/Qualifiers
1. .119733
/organism="Homo sapiens"

```

```
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP4-655C5"
/clone_lib="RPC1-4"
repeat_region
1..433
/note="L2 repeat: matches 1703..2138 of consensus"
460..950
/note="L2 repeat: matches 1042..1597 of consensus"
misc_feature
complement(1180..1481)
/note="match: STS: Em:HS168L15S"
repeat_region
1869..2002
/note="L2 repeat: matches 2608..2750 of consensus"
2158..2498
/note="LTR1 repeat: matches 1..365 of consensus"
repeat_region
2774..2978
/note="L2 repeat: matches 1337..1539 of consensus"
3085..3300
/note="MER6 repeat: matches 34..257 of consensus"
3307..3425
/note="LTR1 repeat: matches 1..101 of consensus"
3444..3662
/note="AluJb repeat: matches 1..219 of consensus"
3666..3842
/note="LTR1 repeat: matches 156..341 of consensus"
complement(3839..4322)
/note="match: GSS: Em:A2419683"
misc_feature
complement(4102..4359)
/note="match: GSS: Em:A2639762"
4153..4623
/note="match: GSS: Em:B87991"
4160..4280
/note="match: GSS: Em:A2264436"
4168..4466
/note="match: GSS: Em:AQ008901"
complement(4217..4491)
/note="match: GSS: Em:AQ207441"
complement(4425..4551)
/note="match: GSS: Em:B46380"
4287..4627
/note="match: GSS: Em:A2264436"
4342..4466
/note="match: GSS: Em:A2266010"
complement(4440..4627)
/note="match: GSS: Em:A234123"
4473..4922
/note="match: GSS: Em:A2266010"
4631..4758
/note="MER85 repeat: matches 1..139 of consensus"
complement(4759..5131)
/note="match: GSS: Em:AQ73957"
complement(4759..5141)
/note="match: GSS: Em:B80743"
4840..5087
/note="match: GSS: Em:B81269"
4840..5235
/note="match: GSS: Em:AQ226387"
4995..5171
/note="match: GSS: Em:A2361655"
complement(5054..5532)
/note="match: GSS: Em:AQ636693"
5199..5586
/note="match: GSS: Em:A2361655"
complement(5300..5752)
/note="match: GSS: Em:A246810"
5749..6000
/note="AluJo repeat: matches 1..254 of consensus"
6027..6060
/note="MER76 repeat: matches 1..34 of consensus"
6061..6105
/note="L2 repeat: matches 2645..2700 of consensus"
6113..6160
/note="4 copies 12 mer 83% conserved"
6115..6154
repeat_region

/note="10 copies 4 mer tga 87% conserved"
7753..8033
/note="AluX repeat: matches 1..289 of consensus"
8499..8601
/note="LIM8 repeat: matches 6184..6286 of consensus"
9640..9908
/note="LIM5 repeat: matches 7609..7898 of consensus"
complement(10367..10902)
/note="match: GSS: Em:A2520653"
11002..11211
/note="MIR repeat: matches 31..259 of consensus"
11273..11378
/note="L2 repeat: matches 2543..2658 of consensus"
11465..11648
/note="LTR1 repeat: matches 222..394 of consensus"
12112..12163
/note="26 copies 2 mer tt 71% conserved"
12216..12511
/note="AluSq repeat: matches 1..299 of consensus"
13635..13818
/note="MIR repeat: matches 58..241 of consensus"
14327..14590
/note="3 copies 88 mer 92% conserved"
16291..16453
/note="AluSq/X repeat: matches 135..297 of consensus"
16454..16489
/note="Tigger4(Zombi) repeat: matches 45..80 of consensus"
16518..16607
/note="Tigger4(Zombi) repeat: matches 2642..2731 of consensus"
complement(16856..17240)
/note="match: GSS: Em:AQ245004"
17145..17319
/note="MER3A repeat: matches 1..181 of consensus"
17602..17902
/note="AluSc repeat: matches 2..301 of consensus"
18131..18420
/note="AluX repeat: matches 1..291 of consensus"
19188..19469
/note="AluX repeat: matches 1..297 of consensus"
19497..19762
/note="AluJo repeat: matches 1..302 of consensus"
21340..21430
/note="L2 repeat: matches 2564..2667 of consensus"
21479..21544
/note="2 copies 33 mer 93% conserved"
22600..23022
/note="LTR1A8 repeat: matches 211..608 of consensus"
23011..23058
/note="LTR1A8 repeat: matches 1..46 of consensus"
23126..23881
/note="HERV4 repeat: matches 5030..5801 of consensus"
complement(23734..23982)
/note="match: GSS: Em:AQ462686"
23931..25376
/note="HERV17 repeat: matches 2365..5466 of consensus"
25361..26415
/note="HERV17 repeat: matches 985..1987 of consensus"
26416..26572
/note="Charliel repeat: matches 1..177 of consensus"
26622..26828
/note="Charliel repeat: matches 2506..2757 of consensus"
26850..27019
/note="MIR repeat: matches 23..209 of consensus"
28015..28168
/note="Tigger2a repeat: matches 273..429 of consensus"
28169..28475
/note="AluY repeat: matches 1..311 of consensus"
complement(28448..28942)
/note="match: GSS: Em:AQ663299"
28476..28760
/note="Tigger2a repeat: matches 1..273 of consensus"
29037..29340
repeat_region
```

```

misc_feature      /note="AluX repeat: matches 1..303 of consensus"
                  31526..33561
                  /note="Cpg island"
                  /evidence=not_experimental
repeat_region     36658..36962
                  /note="Alu repeat: matches 1..304 of consensus"
                  repeat_region 37387..37481

Query Match      64.0%; Score 16; DB 9; Length 119733;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 aagtcacgaattatt 24
      |||
Db 3994 AAGTCACAGATTATT 3979

RESULT 41
AC008951 124103 bp DNA HTG 01-JUL-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2337A12, WORKING DRAFT
ACCESSION AC008951
VERSION AC008951.5 GI:14579694
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 124103)
JOURNAL DOE Joint Genome Institute.
REFERENCE Sequencing of Human Chromosome 5
AUTHORS Unpublished
TITLE 2 (bases 1 to 124103)
JOURNAL DOE Joint Genome Institute.
REFERENCE Direct Submission.
AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL On Jul 1, 2001 this sequence version replaced gi:7710864.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 756969
Center clone name: CTRB-H1_2337A12
-----
Summary Statistics
Consensus quality: 120091 bases at least Q40
Consensus quality: 122780 bases at least Q30
Consensus quality: 123423 bases at least Q20
Estimated insert size: 123000; pulse field gel estimation
Estimated insert size: 123803; sum-of-contrigs estimation
Quality coverage: 5.58 in Q20 bases; pulse field gel estimation
Quality coverage: 5.54 in Q20 bases; sum-of-contrigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contrigs. Gaps between the contrigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given. However the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 29001: contrig of 29001 bp in length
* 29002 29101: gap of unknown length
* 29102 48451: contrig of 19350 bp in length
* 48452 48551: gap of unknown length
* 48552 51607: contrig of 3056 bp in length
* 51608 51707: gap of unknown length
* 51708 124103: contrig of 72396 bp in length.
Location/Qualifiers
FEATURES
1..129010
/organism="Homo sapiens"

```

```

source
1..124103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2337A12"
BASE COUNT 34654 a 24183 c 24436 g 40530 t 300 others
ORIGIN

Query Match      64.0%; Score 16; DB 2; Length 124103;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 agtcacgaattatt 25
      |||
Db 24194 AGTCACAGATTATT 24179

RESULT 42
AL159168
LOCUS AL159168
DEFINITION Human DNA sequence from clone RP11-401H23 on chromosome 9, complete
ACCESSION AL159168
VERSION AL159168.15 GI:10443391
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 129010)
JOURNAL Kimberley A.
REFERENCE Direct Submission
AUTHORS Submitted (26-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
TITLE CB10 ISA, UK. E-mail enquiries: humquerry@sanger.ac.uk
JOURNAL requests: clonerequest@sanger.ac.uk
COMMENT On Oct 1, 2000 this sequence version replaced gi:10178449.
during sequence assembly data is compared from overlapping clones.
where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contrigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-401H23 is from the library RCT-11.2 constructed at the
Roswell Park Cancer Institute by the group of Pieter de Jong. For
further details see http://bacpac.med.buffalo.edu/
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-401H23 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-401H23 is at 1 in this sequence.
The true left end of clone RP11-16121 is at 128911 in this
sequence. The true right end of clone RP11-38608 is at 19333 in
this sequence.
Location/Qualifiers
1..129010
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/clone="RP11-401H23"
/clone_lib="RPC1-11.2"
BASE COUNT      39045 a 27222 c 2654 g 36089 t
ORIGIN

```

```

Query Match      64.0% Score 16: DB 9: Length 129010;
Best Local Similarity 100.0%; Pred. No. 13:
Matches 16: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      10 agtcacagaattactt 25
|||||
Db 102456 AGTCACAGATTATTT 102471

```

```

RESULT 43
AC019289/c
LOCUS
DEFINITION Homo sapiens chromosome 4 clone RP11-203119 map 4, WORKING DRAFT
SEQUENCE, 12 unordered pieces.
AC019289
AC019289.3 GI:7528059
HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Bara,N., Beckert,R., Bede,F.,
Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G., Caste,A.,
Chapel,V., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fenebor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,S., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
Mcneeters,R., Meldrum,J., Menues,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Piette,N., Plisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody.M.

```

```

TITLE
JOURNAL Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 8, 2000 this sequence version replaced gi:6778536.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4306
Center clone name: 203_1_19
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 13255 bases at least Q40
Consensus quality: 13586 bases at least Q30
Consensus quality: 137053 bases at least Q20

```

```

Insert size: 140000; agarose-fp
Insert size: 138118; sum-of-contigs
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-contigs
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1      879: contig of 879 bp in length
*      880 979: gap of 100 bp
*      980 2419: contig of 1340 bp in length
*      2420 2419: gap of 100 bp
*      2420 6972: contig of 4553 bp in length
*      6973 7072: gap of 100 bp
*      7073 11233: contig of 4161 bp in length
*      11234 11333: gap of 100 bp
*      11334 18294: contig of 6961 bp in length
*      18295 18394: gap of 100 bp
*      18395 24919: contig of 6525 bp in length
*      24920 25019: gap of 100 bp
*      25020 32971: contig of 7952 bp in length
*      32972 33071: gap of 100 bp
*      33072 44845: contig of 11774 bp in length
*      44846 44945: gap of 100 bp
*      44946 60683: contig of 15738 bp in length
*      60684 60783: gap of 100 bp
*      60784 77274: contig of 16491 bp in length
*      77275 77374: gap of 100 bp
*      77375 102352: contig of 24978 bp in length
*      102353 102452: gap of 100 bp
*      102453 139218: contig of 36766 bp in length.
FEATURES
source
1. 139218
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-203119"
/clone_lib="RPC1-11 Human Male BAC"
1. 879
/feature="assembly-fragment"
clone_end:SP6
vector_side:left"
980..2319
/feature="assembly-fragment"
clone_end:T7
vector_side:right"
2420..6972
/feature="assembly-fragment"
7073..11233
/feature="assembly-fragment"
11334..18294
/feature="assembly-fragment"
18395..24919
/feature="assembly-fragment"
25020..32971
/feature="assembly-fragment"
33072..44845
/feature="assembly-fragment"
44946..60683
/feature="assembly-fragment"
60784..77274
/feature="assembly-fragment"
77375..102352
/feature="assembly-fragment"
102453..139218
/feature="assembly-fragment"

```

```

misc_feature 1. 879
/feature="assembly-fragment"
clone_end:SP6
vector_side:left"
980..2319
/feature="assembly-fragment"
clone_end:T7
vector_side:right"
2420..6972
/feature="assembly-fragment"
7073..11233
/feature="assembly-fragment"
11334..18294
/feature="assembly-fragment"
18395..24919
/feature="assembly-fragment"
25020..32971
/feature="assembly-fragment"
33072..44845
/feature="assembly-fragment"
44946..60683
/feature="assembly-fragment"
60784..77274
/feature="assembly-fragment"
77375..102352
/feature="assembly-fragment"
102453..139218
/feature="assembly-fragment"
misc_feature 43319 a 25775 c 25817 g 43205 t 1102 others
BASE COUNT

```

ORIGIN

Query Match 64.0%; Score 16; DB 2; Length 139218;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 agtcacagaattatt 25
 |||||||

Db 103832 AGTCACAGATTATT 103817

RESULT 44
 AC092893/C
 LOCUS

DEFINITION Homo sapiens chromosome 3q clone RP11-203119, WORKING DRAFT

AC092893 140002 bp DNA

AC092893 SEQUENCE, 9 unordered pieces.

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarta,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Boucek,J., Bowler,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Huylk,S., Hune,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryak,J., Kovari,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisgeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Nektson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S., Ogund,M., Okunolu,G., Ofaquye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Ruiz,S., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Sison,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swalek,A., Tabor,P., Tamejisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Verba,V., Villalón,D., Vinton,R., Wall,R., Wang,S., Ward-Moore,S., Warren,C., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

COMMENT

Direct Submission
 Unpublished
 2 (bases 1 to 140002)
 Worley,K.C.
 Direct Submission
 Submitted (09-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HCSH
 Center clone name: RP11-203119

----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 134207 bases at least Q40
 Consensus quality: 137098 bases at least Q30
 Consensus quality: 138384 bases at least Q20
 Estimated insert size: 138999; sum-of-contigs estimation
 quality coverage: 0x in Q20 bases; agarose-fp estimation
 quality coverage: 4.8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 9 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 38256: contig of 38256 bp in length
 * 38257 38356: gap of unknown length
 * 38357 61565: contig of 23209 bp in length
 * 61566 61665: gap of unknown length
 * 61666 78299: contig of 16634 bp in length
 * 78300 78399: gap of unknown length
 * 78400 93197: contig of 14798 bp in length
 * 93198 93297: gap of unknown length
 * 93298 109082: contig of 15785 bp in length
 * 109083 109182: gap of unknown length
 * 109183 120959: contig of 11777 bp in length
 * 120960 121059: gap of unknown length
 * 121060 128609: contig of 7550 bp in length
 * 128610 128709: gap of unknown length
 * 128710 135281: contig of 6572 bp in length
 * 135282 135381: gap of unknown length
 * 135382 140002: contig of 4621 bp in length.

FEATURES

Source

1. 140002
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3q"
 /clone="RP11-203119"
 BASE COUNT 44168 a 26034 c 26002 g 42981 t 817 others

ORIGIN

Query Match 64.0%; Score 16; DB 2; Length 140002;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 agtcacagaattatt 25
 |||||||

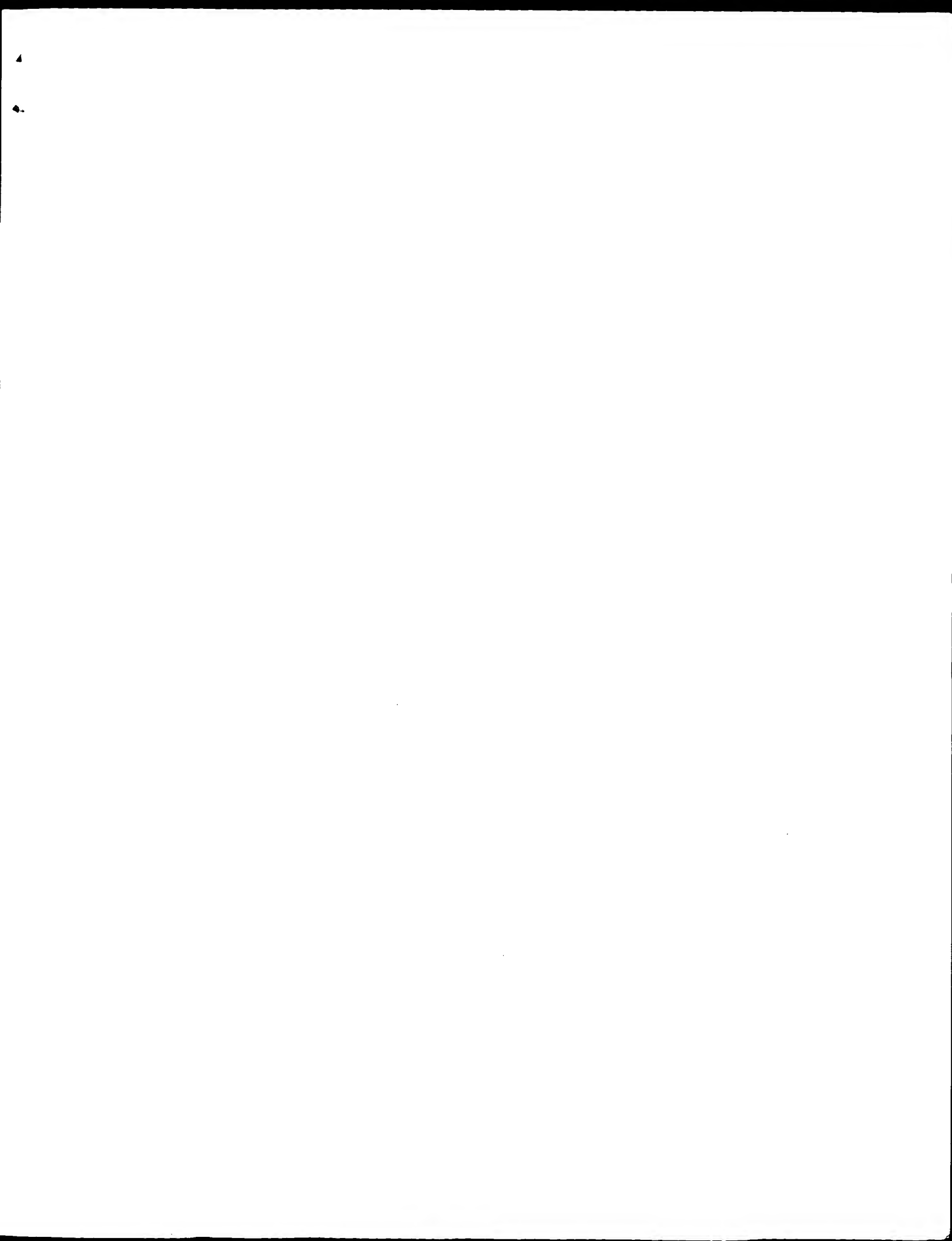
Db 1767 AGTCACAGATTATT 1752

RESULT 45
 AC093258

LOCUS AC093258 151206 bp DNA 16-AUG-2001
 DEFINITION Homo sapiens chromosome 5 clone RP11-1E3, WORKING DRAFT SEQUENCE, 1
 AC093258 unordered pieces.

AC093258 AC093258.1 GI:15193392
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:16:28 : Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196f-2

Perfect score: 25

Sequence: 1 atgttcgaagtcacagaattatt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 39280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : N_Geneseq_1101.*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA013303 E. coli B101 syn
3	25	100.0	1084	10	AA013329 E. coli B101 syn
4	25	100.0	1121	7	AA060496 Sequence encoding
5	25	100.0	5872	15	AA062386 Biotin-biosynthesis
6	17	68.0	17	14	AA056307 Biotin DNA primer
7	15	60.0	561	22	AA094235 primer specific fo
8	15	60.0	1771	20	AA230602 Mouse integrin alp
9	15	60.0	1771	22	AA010264 Mouse integrin alp
10	15	60.0	3210	22	AA018561 Human cDNA sequence
11	15	60.0	3520	22	AA093909 Human cDNA encoding

C 12	15	60.0	3520	22	AAE29361 Amyloid-beta prote
C 13	15	60.0	3748	22	AAE98729 Human late stage o
C 14	15	60.0	6087	22	AA022683 Human cDNA encodin
C 15	15	60.0	6116	22	AA022447 Human cDNA encodin
C 16	15	60.0	116624	19	AAV52850 Human eyal gene co
C 17	14	56.0	329	21	AAA41925 Human secreted exp
C 18	14	56.0	360	22	AA065421 Novel human polynu
C 19	14	56.0	385	14	AA039649 Expressed Sequence
C 20	14	56.0	385	14	AA059601 Human brain Expres
C 21	14	56.0	403	22	AA113743 Probe #3676 for ge
C 22	14	56.0	403	22	AA135103 Probe #3789 used t
C 23	14	56.0	523	22	AA103624 Probe #3615 used t
C 24	14	56.0	523	22	AAH12853 Human cDNA clone (
C 25	14	56.0	574	22	AA141094 Probe #9780 used t
C 26	14	56.0	902	21	AA256756 Human transmembran
C 27	14	56.0	1050	9	AA080927 Lividans Exported
C 28	14	56.0	1465	22	AAH14446 Human cDNA sequenc
C 29	14	56.0	1677	17	AA031932 Retinoid X recepto
C 30	14	56.0	1787	17	AA031928 Retinoid X recepto
C 31	14	56.0	1890	20	AA252940 Human prostate tum
C 32	14	56.0	1919	22	AA160385 Human polynucleoti
C 33	14	56.0	2022	22	AAH15029 Human cDNA sequenc
C 34	14	56.0	2102	22	AA158599 Human polynucleoti
C 35	14	56.0	2220	22	AA012786 Human nuclear horm
C 36	14	56.0	2233	19	AAV10120 Human retinoid rec
C 37	14	56.0	2585	22	AAE77541 Caenorhabditis bri
C 38	14	56.0	2845	21	AAE69462 Human secreted pro
C 39	14	56.0	3413	20	AAV72330 F. balustinum CP70
C 40	14	56.0	5662	22	AA161089 Human polynucleoti
C 41	14	56.0	5684	21	AA059953 Human secreted pro
C 42	14	56.0	5773	21	AA076014 Human ORFX ORF1569
C 43	14	56.0	5876	22	AA159303 Human polynucleoti
C 44	14	56.0	7063	22	AAH24651 Nucleotide sequenc
C 45	14	56.0	8041	17	AA010105 Adherence confetti
C 46	14	56.0	10126	22	AA163610 Human kidney relat
C 47	14	56.0	14273	22	AA162733 Human breast or ov
C 48	14	56.0	32351	21	AAE21307 Human low adenosin
C 49	14	56.0	32351	21	AAA35185 Human adenosine re
C 50	14	56.0	40298	21	AAE21311 Human low adenosin
C 51	14	56.0	40298	21	AAA35189 Human adenosine re
C 52	14	56.0	81001	22	AA030035 Human apolipoprote
C 53	14	56.0	1038602	20	AA0201425 Complete genome se
C 54	13	52.0	87	16	AA022576 Human gene signatu
C 55	13	52.0	105	16	AA020380 Human gene signatu
C 56	13	52.0	151	22	AA126511 Probe #16444 for g
C 57	13	52.0	151	22	AA155262 Human gene #23948 used
C 58	13	52.0	167	16	AA020233 Human gene signatu
C 59	13	52.0	187	21	AA026213 Human secreted pro
C 60	13	52.0	259	22	AA127363 Probe #17296 for g
C 61	13	52.0	259	22	AA156259 Probe #24945 used
C 62	13	52.0	302	22	AA123024 Probe #12957 used
C 63	13	52.0	302	22	AA148324 Probe #17010 used
C 64	13	52.0	302	22	AA108683 Probe #8674 used t
C 65	13	52.0	335	22	AA125914 Probe #15847 for g
C 66	13	52.0	335	22	AA153184 Probe #21870 used
C 67	13	52.0	330	22	AA035496 Human colon cancer
C 68	13	52.0	343	22	AA163192 Human kidney relat
C 69	13	52.0	354	15	AA077082 Human genome fragm
C 70	13	52.0	363	21	AA050159 Antibody 5H7 heavy
C 71	13	52.0	378	13	AA025667 Sequence of the an
C 72	13	52.0	387	21	AA079285 Human lung tumour
C 73	13	52.0	408	21	AA074604 Murine 5B3 antibod
C 74	13	52.0	409	21	AA038366 Human secreted pro
C 75	13	52.0	415	21	AA056783 Eucalyptus grandis
C 76	13	52.0	441	21	AA007649 Human secreted pro
C 77	13	52.0	476	21	AA007951 Human secreted pro
C 78	13	52.0	498	22	AA009392 Human secreted pro
C 79	13	52.0	498	22	AA050926 Human cyclophilin
C 80	13	52.0	513	22	AA165611 Probe #16561 for g
C 81	13	52.0	513	22	AA155392 Probe #24078 used
C 82	13	52.0	516	22	AA120799 Probe #10732 for g
C 83	13	52.0	516	22	AA146029 Probe #14715 used
C 84	13	52.0	516	22	AA106507 Probe #6498 used t

ALIGNMENTS

```

85 13 52.0 520 22 AA18076 probe #809 for ge
86 13 52.0 520 22 AA13076 probe #11762 used
87 13 52.0 528 22 AA18231 probe #8164 for ge
88 13 52.0 528 22 AA143274 probe #11960 used
89 13 52.0 550 21 AAC94878 Cat flea hindgut a
90 13 52.0 556 22 AAH09733 Human cDNA clone (
91 13 52.0 556 22 AAH10417 Human cDNA clone (
92 13 52.0 559 22 AAF93531 cDNA encoding SRT
93 13 52.0 566 21 AAC07952 Human secreted pro
94 13 52.0 580 22 AA16719 probe #6652 for ge
95 13 52.0 580 22 AA140105 probe #8791 used t
96 13 52.0 585 22 AA117412 probe #7345 for ge
97 13 52.0 585 22 AA142313 probe #10999 used
98 13 52.0 603 22 AAH32562 Human secreted pro
99 13 52.0 607 21 AA16367 Human colon cancer
100 13 52.0 626 21 AA16245 Human colon cancer

RESULT 1
ID AA162941 standard; DNA: 839 BP.
AC AA162941;
XX
XX 22-OCT-2001 (first entry)
DE Human genomic DNA SEQ ID NO 269.
XX
XX Human: noctropic; neuroprotective; cytostatic; dermatological; viroicide;
XX immunosuppressive; antinflammatory; anti-HIV; antibacterial; vulnerary;
XX antiparkinsonian; antisticking; antianaemic; antitarrhritic; cancer;
XX antirheumatic; hepatotropic; cerebroprotective; antinflammatory;
XX antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine;
XX ds.
XX Homo sapiens.
XX
XX WO20015449-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01346.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUL-2000; 2000US-0216880.
XX 14-JUL-2000; 2000US-0218290.
XX 14-AUG-2000; 2000US-0225447.
XX 01-SEP-2000; 2000US-0229343.
XX 06-SEP-2000; 2000US-0230437.
XX 08-SEP-2000; 2000US-0231243.
XX 25-SEP-2000; 2000US-0234997.
XX 29-SEP-2000; 2000US-0236367.
XX 13-OCT-2000; 2000US-0239937.
XX 08-NOV-2000; 2000US-0246476.
XX 08-NOV-2000; 2000US-0246477.
XX 08-NOV-2000; 2000US-0246525.
XX 08-NOV-2000; 2000US-0246526.
XX 08-NOV-2000; 2000US-0246528.
XX 17-NOV-2000; 2000US-0249210.
XX 17-NOV-2000; 2000US-0249211.
XX 17-NOV-2000; 2000US-0249214.
XX 17-NOV-2000; 2000US-0249265.
XX 01-DEC-2000; 2000US-0250160.
XX 01-DEC-2000; 2000US-0250391.
XX 05-DEC-2000; 2000US-0251030.
XX 05-DEC-2000; 2000US-0251988.

```

```

PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WEL; 2001-476225/51.
XX
XX Novel plasma membrane associated proteins useful for diagnosing,
XX treating, preventing and/or prognosing disorders related to the
XX proteins, including cancer, immune response and neuronal disorders
XX
XX Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AA162752-AA162961) and proteins
XX (AA162347-AA162415) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;
XX
XX Query Match 100.0%; Score 25; DB 22; Length 839;
XX Best Local Similarity 100.0%; Pred. No. 1.8e-05;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 atgtgcgaagtcacagaattatt 25
XX 56 atctgcgaagtcacagaattatt 80
XX
XX RESULT 2
XX ID AAX01303 standard; DNA: 1041 BP.
XX
XX AAX01303;
XX
XX 12-APR-1999 (first entry)
XX
XX E. coli biotin synthetase (Biot) coding sequence.
XX
XX DAP aminotransferase; diaminopelargonic acid; transgenic plant;
XX biotin synthase; biotin production; vitamin H; Biot; ss.
XX
XX Escherichia coli.
XX
XX US5869719-A.
XX
XX 09-FEB-1999.
XX
XX 30-APR-1997; 97US-0846338.
XX
XX 30-APR-1997; 97US-0846338.
XX
XX 08-MAR-1995; 95US-0401068.
XX

```

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Patton DA:
 XX
 XX WPI: 1999-152902/13.
 DR P-PSDB: AAW73906.
 XX
 XX Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di:amino-pelargonic acid amino-transferase or biotin
 PT synthase
 XX
 PS Example 2: Column 37-40: 34pp: English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioH). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SQ Sequence 1041 BP: 262 A; 273 C; 305 G; 201 T; 0 other:

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgtgcgaagtcacagaattattt 25
 ||||||||||||||||||||
 DB 24 atgtgcgaagtcacagaattattt 48

RESULT 3

AA091329
 ID AAN91329 standard; DNA: 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 DE E.coli Bio B gene.
 XX
 DE E.coli Bio B gene; biotin.
 XX
 KM E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 XX
 PN GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 XX (UKAG-) UK MIN. AGRIC. FISH.
 PA Pearson BM, McKee RA;
 PI
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3: page 33-4: 52pp: English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 SQ Sequence 1084 BP: 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgtgcgaagtcacagaattattt 25
 ||||||||||||||||||||
 DB 47 atgtgcgaagtcacagaattattt 71

RESULT 4

AA060496
 ID AAN60496 standard; DNA: 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 DE Sequence encoding biotin synthesising enzyme.
 XX
 DE Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
 XX
 KM Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag= a
 XX
 PN JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB: AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure: Page 534; 23pp: Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.
 CC
 SQ Sequence 1121 BP: 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 atgtgcgaagtcacagaattattt 25
 ||||||||||||||||||||
 DB 65 atgtgcgaagtcacagaattattt 89

RESULT 5

AA062386
 ID AAO62386 standard; DNA: 5872 BP.
 XX

XX 13-APR-1994 (first entry)
 DT
 XX
 DE BioB DNA primer.
 XX
 KW BioA; BioB; promoter; biotin; operon; primer; ss.
 XX
 OS Synthetic.
 XX
 PN JP05219956-A.
 XX
 PD 31-AUG-1993.
 XX
 PF 14-SEP-1992; 92JP-0244792.
 XX
 PR 14-SEP-1992; 92JP-0244792.
 XX
 PA (SHIS) SHISEIDO CO LTD.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 DR WPI; 1993-308323/39.
 XX
 PT DNA sequence of biotin operon - has base sequence of E. coli
 PT mutated by base pair(s) compared to wild type
 XX
 PS Example 1: Fig 8; 11pp: Japanese.
 XX
 CC A novel DNA sequence comprises the E.coli biotin operon (BO) in which
 CC the control region of BO or the region near the bioB initiation
 CC codon is mutated by at least one base pair compared to its
 CC wild type. Two primers (AA056306-056307) are described in Example 1.
 CC A microorganism belonging to Escherichia genus, transformed by
 CC a recombinant plasmid carrying such DNA can be used for the prodn.
 CC of biotin-active substances.
 CC
 SQ Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

Query Match 68.0%; Score 17; DB 14; Length 17;
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gtcgaagtcacagaat 20
 |||||
 DB 17 GTGCGAAGTCACAGANT 1

RESULT 7
 AAF94235
 ID AAF94235 standard; DNA; 561 BP.
 XX
 AC AAF94235;
 XX
 DT 23-MAY-2001 (first entry)
 XX
 DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 669.
 XX
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes; PCK primer; ss.
 XX
 OS Synthetic.
 XX
 PN EP1067182-A2.
 XX
 PD 10-JAN-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114090.
 XX
 PR 08-JUL-1999; 99JP-0194179.
 PR 11-JAN-2000; 2000JP-0118775.
 PR 02-MAY-2000; 2000JP-0183766.
 XX
 PA (HELI-) HELIX RES INST.

XX
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX
 DR WPI; 2001-093989/11.
 XX
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -
 XX
 PS Claim 5; SEQ ID 669; 609pp + CD ROM; English.
 XX
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 SQ Sequence 561 BP; 213 A; 83 C; 114 G; 149 T; 2 other;

Query Match 60.0%; Score 15; DB 22; Length 561;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaat 22
 |||||
 DB 339 caagtcacagaat 353

RESULT 8
 AA230602
 ID AA230602 standard; DNA; 1771 BP.
 XX
 AC AA230602;
 XX
 DT 18-JAN-2000 (first entry)
 XX
 DE Mouse integrin alpha 4 coding sequence.
 XX
 KW Human; integrin; antisense; oligonucleotide; inhibition; expression;
 KW very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;
 KW vascular endothelial cell; vascular endothelium; migration; inflammation;
 KW atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
 KW metastasis; circulatory system; autoimmune disease; Grave's disease;
 KW Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
 XX
 OS Mus sp.
 XX
 PN US5968826-A.
 XX
 PD 19-OCT-1999.
 XX
 PF 05-OCT-1998; 98US-0166203.
 XX
 PR 05-OCT-1998; 98US-0166203.


```

PF 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 18732; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 3210 BP; 876 A; 753 C; 607 G; 974 T; 0 other;
SQ

```

Query Match 60.0%; Score 15; DB 22; Length 3210;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 11 gtacacgaattattc 25
Db 2201 GTACACGAATTATT 2187

```

RESULT 11
 AAF93909/C
 ID AAF93909 standard; cDNA: 3520 BP.
 AC AAF93909;
 XX
 DT 23-MAY-2001 (first entry)
 XX
 DE Human cDNA encoding a membrane or secretory protein clone PSEC0256.
 XX
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 XX rheumatoid arthritis; diabetes; ss.
 OS Homo sapiens.
 XX
 PN EP1067182-A2.

```

XX
XX 10-JAN-2001.
PD
XX
XX 07-JUL-2000; 2000EP-0114090.
PF
XX
XX 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI P-SDB; AAB88482.
DR WPI: 2001-093989/11.
XX
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
PT
XX
XX Claim 1; SEQ ID 331; 609pp + CD ROM; English.
XX
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantify the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies (agonists and antagonists) may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
CC (ELISA)). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
XX Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;
SQ

```

Query Match 60.0%; Score 15; DB 22; Length 3520;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 8 caagtcacagatata 22
Db 3182 CAAGTCACAGATTA 3168

```

RESULT 12
 AAF29361/C
 ID AAF29361 standard; DNA: 3520 BP.
 AC AAF29361;
 XX
 DT 20-APR-2001 (first entry)
 XX
 DE Amyloid-beta protein aggregation regulating factor DNA SEQ ID 9.
 XX
 KW Human; amyloid-beta protein; aggregation regulating factor;
 XX Alzheimer's disease; ds.
 OS Homo sapiens.
 XX
 PN

PN WO200104299-A1.
 XX
 PD 18-JAN-2001.
 XX
 PF 06-JUL-2000; 2000WO-JP04515.
 XX
 PR 08-JUL-1999; 99JP-0194179.
 XX 18-OCT-1999; 99US-0159586.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yamazaki M, Satoh S;
 PI Arakawa H, Morita M;
 XX
 DR WPI: 2001-138347/14.
 DR P-PSDB: AAB49771.
 XX
 PT polynucleotide encoding Amyloid-beta protein agglutination-controlling
 PT factor, useful for inhibiting or promoting agglutination or
 PT sedimentation of amyloid-beta protein and in diagnosis and screening
 PT drugs for Alzheimer's disease -
 XX
 PS Claim 1; Page 58-63; 72pp; Japanese.
 XX
 CC This invention relates to polynucleotides AAF29357 - AAF29361 which
 CC encode proteins AAB49767 - AAB49771. The proteins inhibit or promote the
 CC agglutination of amyloid beta protein. The proteins and polynucleotide
 CC sequences are useful in the diagnosis of Alzheimer's disease. They are
 CC also useful for screening drugs which are useful for treating Alzheimer's
 CC disease.
 XX
 SQ Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;
 XX
 Query Match 60.0%; Score 15; DB 22; Length 3520;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 8 caagtcacagaatta 22
 |||||
 DB 3182 CAAGTCACAGAAATTA 3168
 XX
 RESULT 13
 AAF98729/C
 ID AAF98729 standard; DNA; 3748 BP.
 XX
 AC AAF98729;
 XX
 DT 02-JUL-2001 (first entry)
 XX
 DE Human late stage ovarian tumour polynucleotide marker 44.
 XX
 KW Human; ovarian cancer; identification; detection; characterisation;
 KW tumour; kinase; marker; cytostatic; anticense gene therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200118542-A2.
 XX
 PD 15-MAR-2001.
 XX
 PF 01-SEP-2000; 2000WO-US24199.
 XX
 PR 03-SEP-1999; 99US-0152547.
 PR 16-MAR-2000; 2000US-0190347.
 PR 21-MAR-2000; 2000US-0191321.
 PR 31-MAY-2000; 2000US-0208382.
 PR 20-JUL-2000; 2000US-0220467.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Lee J, Thompson P, Lillie J;

XX
 DR WPI: 2001-211428/21.
 XX
 PT Detection, assessment, prevention and therapy of ovarian cancer.
 PT comprises detecting changes in the expression of a variety of markers -
 XX
 PS Claim 1; Page 1196-1197; 1198pp; English.
 XX
 CC The present invention describes a method for assessing whether a patient
 CC is afflicted with ovarian cancer by comparing: (1) the expression of a
 CC marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the
 CC normal level of expression of (I) in a control non-ovarian cancer
 CC sample, where a significant difference between the level of expression
 CC in (a) and (b) is an indication that the patient is afflicted with
 CC ovarian cancer. (1) have cytostatic activities and can be used in
 CC anticense gene therapy. The method, compositions and kits from the
 CC present invention can be used for: (1) assessing and treating ovarian
 CC cancer; (2) making isolated hybridoma, which produces an antibody useful
 CC for ovarian cancer assessment; and (3) inhibiting ovarian cancer in a
 CC patient. AAF98573 to AAF98593 represent human kinase marker primers and
 CC probes which are used in the exemplification of the present invention.
 XX
 SQ Sequence 3748 BP; 1110 A; 825 C; 761 G; 1052 T; 0 other;
 XX
 Query Match 60.0%; Score 15; DB 22; Length 3748;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 8 caagtcacagaatta 22
 |||||
 DB 2627 CAAGTCACAGAAATTA 2613
 XX
 RESULT 14
 AAS22683
 ID AAS22683 standard; CDNA; 6087 BP.
 XX
 AC AAS22683;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Human cDNA encoding a novel human protein #249.
 XX
 KW Human; novel protein; ss; Antinaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
 KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiastrmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200155437-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02623.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YF, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-451939/48.
 DR P-PSDB: AAU14378.
 XX
 PT Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage -
 XX
 SQ Claim 1; Page 676-677; 894pp; English.

XX The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/ elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence encodes a protein of the invention.

XX Sequence 6087 BP; 1490 A; 1609 C; 1574 G; 1413 T; 1 other;

Query Match 60.0%; Score 15; DB 22; Length 6087;

Best Local Similarity 100.0%; Pred. No. 9.4;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 tcgcaagtcacagaa 19

|||||

Db 5191 tcgcaagtcacagaa 5205

RESULT 15

AAS22447 standard; CDNA; 6116 BP.

XX AAS22447;

DT 24-OCT-2001 (first entry)

XX Human CDNA encoding a novel human protein #13.

XX Human: novel protein; ss: Antianemic; osteopathic; antiinflammatory;
XX immunomodulatory; cytosstatic; neuroprotective; vulnerrary; nootropic;
XX anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
XX antibacterial; antiallergic; dermatological; haemostatic; antisthmatic;
XX thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
XX Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
XX tissue regeneration; immune disorder.

XX Homo sapiens.

XX WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-451939/48.

XX P-PSDB; AAU14142.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX Claim 1; Page 190-196; 894pp; English.

XX The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/ elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence encodes a protein of the invention.

XX Sequence 6116 BP; 1491 A; 1626 C; 1582 G; 1417 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 6116;

Best Local Similarity 100.0%; Pred. No. 9.4;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 tcgcaagtcacagaa 19

|||||

Db 5220 tcgcaagtcacagaa 5234

RESULT 16

AAV52850/c standard; DNA; 116624 BP.

XX AAV52850;

DT 06-NOV-1998 (first entry)

XX Human eyal gene contig 4405-9480.

XX Human: eyal; branchio-oto-renal syndrome; BOR; carcinogenesis; cancer;
XX kidney tumourogenesis; tumour; diagnosis; ss.

XX Homo sapiens.

XX WO9832849-A2.

XX 30-JUL-1998.

XX 28-JAN-1998; 98WO-EP00433.

XX 29-JAN-1997; 97US-0036579.

XX (INSP) INST PASTEUR.

XX Abdelhak S, Compain S, Petit C, Vasiliki K, Vincent C;

XX WPI: 1998-427945/36.

XX Nucleic acid corresponding to human genes implicated in
PT branchio-oto-renal syndrome - useful for, e.g. diagnosis and
PT treatment of syndrome and possibly some carcinogenic processes,
PT particularly in kidneys
PS
PS Claim 1; Page 44-105; 191pp: English.

CC The present sequence represents a portion of the human eya-1 gene.
CC Alterations in the eya1 gene are associated with branchio-oto-renal
CC (BOR) syndrome, including the form (BO) without renal anomalies, and
CC possibly in some carcinogenic processes, particularly in the kidney.
CC The related genes eya-2 and eya-3 may also be involved in development
CC of tumours and cell differentiation. Antibodies specific for EYA
CC proteins are used to detect the proteins by immunoassay, while genetic
CC alterations linked to BOR are detected by amplifying DNA with primers
CC (see AAV52857 to AAV52893) then identifying any mutations or deletions.
CC EYA1 and EYA1-B proteins, corresponding nucleic acid, antisense sequences
CC and Ab can be used therapeutically to modulate expression of EYA1(B), or
CC their active derivatives, especially for treating renal disease
CC associated with eyal abnormalities, also for ensuring correct
CC development of grafted organs, especially where these are embryonic.
CC Therapeutic proteins are administered to maintain 0.1-10 mu g/ml in
CC body fluids, locally or systemically. Nucleic acid is administered in
CC usual gene therapy vectors (optionally to cells ex vivo) at doses of
CC 0.1-100 mu g.

SQ Sequence 116624 BP; 35549 A; 20868 C; 23043 G; 37152 T; 12 other;

Query Match 60.0%; Score 15; DB 19; Length 116624;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 gccacagaattattc 25
DB 19735 GTCACGAGATTAATT 19721

RESULT 17
AAAA1925
ID AAAA1925 standard; cDNA: 329 BP.
XX
XX AAAA1925;
XX
XX 21-AUG-2000 (first entry)
XX
XX Human secreted expressed sequence tag SEQ ID NO:665.
XX
XX Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytostatic; antiproliferative; antifungal;
XX antiviral; antidiabetic; antitumour; vulnary; antiparkinsonian;
XX anticancer; osteoprotective; neuroprotective; mototropic; antipsoriatic;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX Parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX tumour; infection; depression; psoriasis; ss.
XX
XX Homo sapiens.
XX
XX WO200021990-A1.
XX
XX 20-APR-2000.
XX
XX 15-OCT-1999; 99WO-US24205.
XX
XX 15-OCT-1998; 98US-0104435.
XX
XX

XX (GENV) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M;
XX
XX WPI; 2000-317937/27.
XX
XX Claim 1; Page 308; 618pp: English.

CC AAA41261 to AAA43419 represent specifically claimed expressed
CC sequence tags (SESTs). Isolated from human, mouse, xenopus and rat
CC tissue sources. The SESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antitumour; vulnary; antiparkinsonian; antipsoriatic; neuroprotective;
CC nocotropic; antiparkinsonian; antiproliferative; cerebroprotective;
CC anticonvulsant; and antidepressant. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA44425 represent linker variants which are given
CC in the exemplification of the present invention.

SQ Sequence 329 BP; 97 A; 84 C; 67 G; 81 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 329;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagaattattc 25
DB 198 tcacagaattattc 211

RESULT 18
AAF65421
ID AAF65421 standard; cDNA: 360 BP.
XX
XX AAF65421;
XX
XX 09-APR-2001 (first entry)
XX
XX Novel human polynucleotide, SEQ ID NO: 1177.
XX
XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;
XX breast cancer; lung cancer; cancer detection; ss.
XX
XX Homo sapiens.
XX
XX WO200102568-A2.
XX
XX 11-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18374.
XX
XX 02-JUL-1999; 99US-0142310.
XX
XX 02-JUL-1999; 99US-0142311.
XX
XX

```

XX (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Chtenjakov R, Drmanac S, Dickson M, Labat I, Leschkowitz D;
PI Kita D, Garcia V, Jones LM, Strache-Grain B;
XX
XX WPI: 2001-091805/10.
XX
XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
XX Claim 9: Page 712; 1046pp; English.
XX
XX The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
XX Sequence 360 BP; 95 A; 70 C; 69 G; 125 T; 1 other:
SQ

```

```

Query Match          56.0%; Score 14; DB 22; Length 360;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 11 gtccagagaattatt 24
   |||||
DB 71 gtccagagaattatt 84

```

```

RESULT 19
AA039649
ID AA039649 standard; DNA; 385 BP.
XX
XX AA039649;
XX
XX 20-MAY-1993 (first entry)
XX
XX Expressed Sequence Tag human gene marker EST000025.
DE
XX expressed sequence tag; human genome project; chromosome;
KW human gene sequencing; PCR mapping; somatic cell hybrids;
KW sublocalisation; gene tagging; tissue typing.
XX
XX Synthetic.
OS
XX
XX MO9300353-A.
XX
XX 07-JAN-1993.
XX
XX 19-JUN-1992; 92WO-US05222.
XX
XX 20-JUN-1991; 91US-0716831.
XX
XX 12-FEB-1992; 92US-0837195.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
XX Adams MD, Venter JC;
PI

```

```

XX
XX WPI: 1993-036325/04.
XX
XX Particular expressed sequence tags from human CDNA - corresponds
PT to transcription prods. of genes, useful for tagging genes,
PT mapping chromosomes and tissue typing
XX
XX Claim 3: Page 86; 199pp; English.
XX
XX This sequence represents an EST (expressed sequence tag) ESTs are
CC markers for human genes actually transcribed in vivo. Unlike the random
CC genomic DNA sequence tagged sites (STSs), ESTs point directly to
CC expressed genes. The use of ESTs could facilitate the tagging of most
CC expressed human genes within a few years at a fraction of the cost of
CC complete genomic sequencing. Using PCR primers AA039419-039580 (sequences
CC designed from the ESTs) sublocalisation of an EST can be achieved with
CC panels of fragments from specific chromosomes or pools of large genomic
CC clones in an analogous manner. This sequence represents EST00025.
XX
XX Sequence 385 BP; 101 A; 66 C; 78 G; 139 T; 1 other:
SQ

```

```

Query Match          56.0%; Score 14; DB 14; Length 385;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 12 tcacagaattatt 25
   |||||
DB 116 tcacagaattatt 129

```

```

RESULT 20
AA059061
ID AA059061 standard; CDNA; 385 BP.
XX
XX AA059061;
XX
XX 16-MAR-1994 (first entry)
XX
XX Human brain Expressed Sequence Tag EST000025.
DE
XX Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
XX Homo sapiens.
OS
XX
XX MO9316178-A.
XX
XX 19-AUG-1993.
XX
XX 12-FEB-1993; 93WO-US01294.
XX
XX 12-FEB-1992; 92US-0837195.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
XX Adams MD, Moreno RF, Venter CJ;
XX
XX WPI: 1993-272882/34.
XX
XX Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
XX
XX Example 1: Page 115; 500pp; English.
XX
XX The Expressed Sequence Tag was isolated from a human brain CDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC

```

CC EST00025 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM and hybridises to chromosome 20.
CC See also AA059041-Q61440.

XX
SQ Sequence 385 BP; 99 A; 66 C; 78 G; 140 T; 2 other;

Query Match 56.0%; Score 14; DB 14; Length 385;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tcacagaattattt 25
Db 116 tcacagaattattt 129

RESULT 21
AA13743
ID AA13743 standard; DNA; 403 BP.

AC AA13743;

DE 12-OCT-2001 (first entry)

XX Probe #3676 for gene expression analysis in human cervical cell sample.

KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.

OS Homo sapiens.

PN W0200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -

PS Claim 25; SEQ ID No 3676; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 403 BP; 107 A; 87 C; 91 G; 118 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 403;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tcacagaattattt 25
Db 162 tcacagaattattt 175

RESULT 22
AA135103
ID AA135103 standard; DNA; 403 BP.

AC AA135103;

DE 17-OCT-2001 (first entry)

XX Probe #3789 used to measure gene expression in human placenta sample.

KW Probe; human; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.

OS Homo sapiens.

PN W0200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488997/53.

PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -

PS Claim 25; SEQ ID No 3789; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX Sequence 403 BP; 107 A; 87 C; 91 G; 118 T; 0 other;

QY 12 tcacagaattattt 25
Db 162 tcacagaattattt 175

RESULT 23
AA103624
ID AA103624 standard; DNA; 403 BP.

AC AA103624;

DE 09-OCT-2001 (first entry)

XX Probe #3615 used to measure gene expression in human breast sample.

XX Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPT: 2001-476286/51.
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 25; SEQ ID No 3615; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocytic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from Wipo
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX Sequence 403 BP; 107 A; 87 C; 91 G; 118 T; 0 other;
 SQ

Query Match 56.0%; Score 14; DB 22; Length 403;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagaattattt 25
 Db 162 tcacagaattattt 175

RESULT 24
 AAH12853
 ID AAH12853 standard; cDNA: 523 BP.
 XX
 AC AAH12853;
 XX
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (3'-primer) SEQ ID NO:9688.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX

PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPT: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX Claim 3; SEQ ID 9688; 2537pp + CD ROM; English.
 PS
 XX
 XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination of
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences. AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 523 BP; 206 A; 87 C; 70 G; 151 T; 9 other;

Query Match 56.0%; Score 14; DB 22; Length 523;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaatt 21
 Db 390 caagtcacagaatt 403

RESULT 25
 AA141094/C
 ID AA141094 standard; DNA: 574 BP.
 XX
 AC AA141094;
 XX
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #9780 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX

```

XX XX W0200157272-A2.
XX XX
XX XX 09-AUG-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US00663.
XX XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
XX DR WPI: 2001-488897/53.
XX XX
XX PR Human genome-derived single exon nucleic acid probes useful for
XX PR analyzing gene expression in human placenta -
XX PS
XX PS Claim 25; SEQ ID NO 9780; 654bp; English.
XX XX
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX XX
XX SQ Sequence 574 BP; 229 A; 109 C; 104 G; 132 T; 0 other;

Query Match          56.0%; Score 14; DB 22; Length 574;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tcacagattttt 25
   |||||||
Db 48 TCACAGATTATT 35

RESULT 26
AAZ56756/c
ID AAZ56756 standard; cDNA; 902 BP.
XX
XX AC AAZ56756;
XX XX
XX DT 23-MAR-2000 (first entry)
XX XX
XX DE Human transmembrane protein HTMPN-59 encoding cDNA.
XX XX
XX KM Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
XX KM antiproliferative; neuroprotective; immune disorder;
XX KM reproductive disorder; smooth muscle disorder; neurological disorder;
XX KM gastrointestinal disorder; developmental disorder;
XX KM cell proliferative disorder; ss.
XX XX
XX OS Homo sapiens.
XX XX
XX PN W09961471-A2.
XX XX
XX PD 02-DEC-1999.
XX XX
XX PF 28-MAY-1999; 99WO-US11904.
XX XX
XX PR 29-MAY-1998; 98US-0087260.
XX PR 02-JUL-1998; 98US-0091674.
XX PR 02-OCT-1998; 98US-0102954.
XX PR 24-NOV-1998; 98US-0109869.
XX XX

```

```

PA (INCY-) INCYTE PHARM INC.
XX XX
XX PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
XX PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
XX PI Au-Young J;
XX XX
XX DR WPI: 2000-072605/06.
XX DR P-PSDB; AAY57935.
XX XX
XX PT Proteins, polynucleotides, vectors, host cells and antibodies used to
XX PT diagnose, treat or prevent immune, reproductive, smooth muscle,
XX PT neurological, gastrointestinal, developmental and cell proliferative
XX PT disorders -
XX XX
XX PS Claim 9; Page 215; 229pp; English.
XX XX
XX CC AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
XX CC transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
XX CC The transmembrane protein have immunospecific, antiproliferative and
XX CC neuroprotective activities. The human transmembrane proteins,
XX CC polynucleotides encoding them and other compositions and methods from
XX CC the present invention, can be used for the diagnosis, treatment or
XX CC prevention of immune, reproductive, smooth muscle, neurological,
XX CC gastrointestinal, developmental and cell proliferative disorders. The
XX CC HTMPN's can be used to treat or prevent disorders associated with a
XX CC decreased expression or activity of HTMPN.
XX XX
XX SQ Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other;

Query Match          56.0%; Score 14; DB 21; Length 902;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagattat 23
   |||||||
Db 425 AGTCACAGATTATT 412

RESULT 27
AAN80927/c
ID AAN80927 standard; DNA; 1050 BP.
XX
XX AC AAN80927;
XX XX
XX DT 15-OCT-1990 (first entry)
XX XX
XX DE Lividans Exported Protein gene.
XX XX
XX KM Longisporus trypsin inhibitor; Lividans exported protein; LTI; LEP-10;
XX KM protease inhibitor; probe; ss.
XX XX
XX OS Streptomyces lividans strain 1326.
XX XX
XX FH Key
XX FH sig_peptide Location/Qualifiers
XX FH FT 394..498
XX FH FT /*tag= a
XX FH FT 499..825
XX FH FT /*tag= b
XX FH FT /product=LEP-10
XX XX
XX PN W08801278-A.
XX XX
XX PD 25-FEB-1988.
XX XX
XX PF 17-AUG-1987; 87WO-US02009.
XX XX
XX PR 18-AUG-1986; 86US-0897245.
XX XX
XX PA (SMIK ) SMITHKLINE BECKMAN CORP.
XX PA (BERK/) TR BERKA.
XX XX
XX PI Berka TR, Formwald JA, Gorniak JG, Rosenberg M;

```


XX WPI: 1988-063996/09.
 DR P-PSDB: AAP80944.
 XX Novel protease inhibitors LEP-10 and LPI - produced by Streptomyces spp.
 PT
 XX Disclosure; ; pp: English.
 XX
 CC The sequence can be cloned and used to produce recombinant LEP-10.
 CC The DNA can be used as a probe for identifying further coding
 CC sequences for protease inhibitors; the protein is useful as an
 CC assay reagent.
 CC See also AAN80926.
 XX
 S0 Sequence 1050 BP; 149 A; 398 C; 366 G; 137 T; 0 other;

Query Match 56.0%; Score 14; DB 9; Length 1050;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 aagtcacagaatta 22
 |||||
 DB 251 AAGTCACAGAAATTA 238

RESULT 28
 AAH14046/c
 ID AAH14046 standard; CDNA: 1465 BP.
 AC
 XX AAH14046:
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:11168.
 XX
 KM Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 11168; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 S0 Sequence 1465 BP; 470 A; 259 C; 270 G; 466 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1465;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaatt 21
 |||||
 DB 1076 CAAGTCACAGAAATT 1063

RESULT 29
 AAT31932/c
 ID AAT31932 standard; DNA: 1677 BP.
 AC
 XX AAT31932:
 DT 27-SEP-1996 (first entry)
 XX
 DE Retinoid X receptor interacting protein R1P14-2 DNA.
 XX
 KM Retinoid X receptor interacting protein; RXR; R1P; R1P14-2; ss.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1356
 FT polyA_signal /*lag= a
 FT 1672-1677 /*tag= b
 FT
 FT WO9621677-A1.
 XX
 PD 18-JUL-1996.
 XX
 PF 08-DEC-1995; 95WO-US16311.
 XX
 PR 13-JAN-1995; 95US-0372652.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Choi H, Moore D, Seol W;
 XX
 DR WPI: 1996-342241/34.
 DR P-PSDB: AAR99739.
 XX
 PT Retinoid X receptor (RXR) interacting protein (RIP) - useful to
 PT modulate or mediate RXR function, anti-RIP antibodies can be used to
 PT determine RIP subcellular distribution patterns
 XX
 PS Claim 8; Page 59-60; 90pp; English.
 XX
 CC A cDNA clone (AAT31932) codes for retinoid X receptor interacting
 CC protein R1P14-2 (AAR99739), a previously undescribed orphan member
 CC of the nuclear receptor superfamily which may be involved in the
 CC complex retinoid response. It was isolated from a mouse liver

CC cDNA library using an RIP14 probe. Different RIP14-2 clones had
 CC different 5' sequences (see also AAT31935-37). Another isoform,
 CC RIP14-1, was also detected (see also AAT31928). The cDNA can be
 CC used for the recombinant prodn. of RIP14-2 in transformed host
 CC cells.

SO Sequence 1677 BP; 486 A; 385 C; 408 G; 398 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 1677;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Oy 2 ttgtcgcaagtcac 15
 |||||
 Db 630 TTGTGCGAAGTCAC 617

RESULT 30

AAT31928/C
 ID AAT31928 standard; DNA: 1787 BP.

AC AAT31928;

DT 27-SEP-1996 (first entry)

DE Retinoid X receptor interacting protein RIP14-1 DNA.

KW Retinoid X receptor interacting protein; RXR; RIP; RIP14-1; ss.

OS Mus sp.

XX Key Location/Qualifiers

FT CDS 12..1466

FT polyA_signal 1782..1787

FT /*tag= b

PN W09621677-A1.

PD 18-JUL-1996.

PF 08-DEC-1995; 95WO-US16311.

PR 13-JAN-1995; 95US-0372652.

PA (GEHO) GEN HOSPITAL CORP.

PI Choi H, Moore D, Seol W;

DR WPI: 1996-342241/34.

DR P-PSDB; AAR99735.

XX Retinoid X receptor (RXR) interacting protein (RIP) - useful to
 XX modulate or mediate RXR function, anti-RIP antibodies can be used to
 XX determine RIP subcellular distribution patterns

XX Claim 8; Page 53-54; 90pp; English.

XX A full-length cDNA clone (AAT31928) codes for retinoid X receptor
 CC interacting protein RIP14-1 (AAR99735), a previously undescribed
 CC orphan member of the nuclear receptor superfamily which may be
 CC involved in the complex retinoid response. It was isolated from
 CC a mouse liver cDNA library using an RIP14 probe. Another isoform,
 CC RIP14-2, was also detected (see also AAT31932). The cDNA can be used
 CC for the recombinant prodn. of RIP14-1 in transformed host cells.

SO Sequence 1787 BP; 515 A; 408 C; 442 G; 422 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 1787;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Oy 2 ttgtcgcaagtcac 15
 |||||
 Db 740 TTGTGCGAAGTCAC 727

RESULT 31

AAZ52940/C
 ID AAZ52940 standard; cDNA: 1890 BP.

AC AAZ52940;

DT 14-MAR-2000 (first entry)

DE Human prostate tumor cDNA library derived EST fragment #83.

KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
 treatment; ds.

OS Homo sapiens.

PN DE19820190-A1.

PD 04-NOV-1999.

PF 28-APR-1998; 98DE-1020190.

PR 28-APR-1998; 98DE-1020190.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPI: 1999-621386/54.

DR P-PSDB; AAY74061, AAY74062, AAY74063.

PT New human nucleic acid sequences from pancreatic tumors, and related
 PT proteins -

PS Claim 2; Page 250; 502pp; German.

XX This invention describes novel polypeptides and their encoding nucleic
 CC acids derived from human pancreatic tumor tissue which have cytostatic
 CC activity. The sequences are also useful in producing pharmaceutical
 CC compositions for treatment of pancreatic tumors. AAZ52838-253014
 CC represent expressed sequence tag (EST) fragments derived from a human
 CC pancreatic tumor cDNA library and which encode the proteins represented
 CC in AAY73814-Y74252.

SO Sequence 1890 BP; 590 A; 358 C; 398 G; 544 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1890;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Oy 8 caagtcacagaatt 21
 |||||

Db 1475 CAAGTCACAGAAATT 1462

RESULT 32

AA160385
 ID AA160385 standard; cDNA: 1919 BP.

AC AA160385;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide seq ID NO 4374.

KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0586042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM41229.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1: SEQ ID NO 4374; 10078pp; English.
 XX
 SS The invention relates to human nucleic acids (AA15798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 1919 BP; 564 A; 394 C; 343 G; 617 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 1919;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caatcacagaatt 21
 ||||||||||||
 DB 402 caagtccagaatt 415

RESULT 33
 AAH15029/C
 ID AAH15029 standard; cDNA: 2022 BP.
 AC AAH15029;
 XX
 DT 26-JUN-2001 (first entry)
 XX

DE Human cDNA sequence SEQ ID NO:13001.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 13001; 2537pp + CD ROM; English.
 XX
 SS The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2022 BP; 622 A; 389 C; 418 G; 593 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2022;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caatcacagaatt 21
 ||||||||||||
 DB 1633 CAAGTCACAGATT 1620

RESULT 34
 AA158599/C
 ID AA158599 standard; cDNA: 2102 BP.
 AC AA158599;
 XX
 DT
 XX

22-OCT-2001 (first entry)

Human polynucleotide SEQ ID NO 802.

Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemolactic; chemokine; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US34263.

21-JAN-2000; 2000US-0486725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

Tang YN, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wang Z, Wehrman T, Xu C, Xue AI, Yang Y, Zhang J; Zhao QA, Zhou P, Goodrich R, Drmanac RT.

WPI: 2001-442253/47.

P-PSDB: AAM39443.

Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

Claim 1; SEQ ID NO 802; 10078pp; English.

The invention relates to human nucleic acids (A157798-A161369) and the encoded polypeptides (AAM38642-AAM42213) with noctropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification.

Sequence 2102 BP; 662 A; 390 C; 441 G; 609 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2102;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

8 caagtcacagatt 21
|||||

Db 1564 CAACTCACAGATT 1551

ADD12786 standard; cDNA; 2220 BP.

ADD12786;

23-OCT-2001 (first entry)

Human nuclear hormone receptor (NHRRC)-3 cDNA.

Human; nuclear hormone receptor; NHRRC-3; intracellular receptor; cancer; transcriptional regulator; metabolic disorder; lysosomal storage disease; Addison's disease; cystic fibrosis; endocrine disorder; atherosclerosis; breast cancer; osteoporosis; Cushing's disease; reproductive disorder; developmental disorder; infertility; renal tubular acidosis; cirrhosis; endometriosis; hereditary neuropathy; immunological disorder; anaemia; acquired immunodeficiency syndrome; AIDS; cell proliferative disorder; allergy; Grave's disease; gene therapy; drug screening; cytostatic; immunomodulatory; gout; diabetes; ss.

Homo sapiens.

Key Location/Qualifiers
CDS 337..1770
/tag=a
/product="Human nuclear hormone receptor (NHRRC)-3"

WO200155392-A2.

02-AUG-2001.

26-JAN-2001; 2001WO-US02741.

28-JAN-2000; 2000US-0178578.

25-FEB-2000; 2000US-0185079.

17-MAR-2000; 2000US-0189999.

(INCY-) INCYTE GENOMICS INC.

Burford N, Baughn MR, Tang YT, Bandman O, Tribouley CM, Das D; Policky JL, Lo TP;

WPI: 2001-488797/53.

P-PSDB: AAE06680.

New human nuclear hormone receptors and polynucleotides encoding the receptors, useful for diagnosing, treating and preventing metabolic, endocrine, reproductive, developmental, immunological and cell proliferative disorders -

Claim 5; Page 95-96; 99pp; English.

The present sequence is human nuclear hormone receptor (NHRRC)-3 cDNA. NHRRC known as nuclear receptor or intracellular receptor, constitute a protein superfamily whose members are both receptors and transcriptional regulators. NHRRC is useful in the diagnosis, treatment and prevention of metabolic (e.g. Addison's disease, cystic fibrosis, diabetes, lysosomal storage disease), endocrine (e.g. disorders of pituitary, thyroid, Cushing's disease, osteoporosis), reproductive (e.g. infertility, endometriosis, breast cancer), developmental (e.g. anaemia, renal tubular acidosis, hereditary neuropathy), immunological disorders (e.g. acquired immunodeficiency syndrome (AIDS), allergy, Grave's disease, gout) and cell proliferative disorders (e.g. atherosclerosis, cirrhosis, cancer) and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of nuclear hormone receptors. NHRRC may also be used to screen for compounds that modulate the activity of NHRRC, for somatic or germline gene therapy, to detect single nucleotide polymorphisms, as element on microarray, to generate hybridisation probes and in drug screening.

Sequence 2220 BP; 710 A; 425 C; 465 G; 620 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2220;
Best Local Similarity 100.0%; Pred. No. 35;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ttgtcgaagtcac 15
 |||||
 Db 1041 TTGTGCAAGTCAC 1028

RESULT 36

AAV10120/C
 ID AAV10120 standard; cDNA; 2233 BP.

XX AAV10120;

XX 29-MAY-1998 (first entry)

XX Human retinoid receptor RRI cDNA.

XX Retinoid receptor; RRI; steroid receptor; agonist; antagonist; cancer;

KW adrenal deficiency; skin disorder; inflammatory disorder;

KW immune response regulator; autoimmune disease; therapeutic antibody; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 363..1781

XX FT /**tag= a

XX FT /product= RRI

XX FT /note= "retinoid receptor"

XX US5728548-A.

XX 17-MAR-1998.

XX 29-JUN-1995; 95US-0496631.

XX 29-JUN-1995; 95US-0496631.

XX (GENY) GENETICS INST INC.

XX Bowman M;

XX WPI: 1998-206567/18.

XX P-PSDB; AAW40072.

XX Human retinoid receptor protein RRI - useful for, e.g. drug

XX screening, therapy and antibody production

XX Claim 8; Column 11-15; 13pp; English.

XX This sequence encodes a novel human steroid receptor, the retinoid

XX receptor protein or RRI. This protein can be used in screening assays

XX for steroid hormone receptor agonists and antagonists and in

XX pharmaceutical compositions for treating adrenal deficiencies, e.g.

XX Addison's disease, cancer, skin disorders, e.g. acne and psoriasis,

XX inflammatory disorders, e.g. arthritis and HIV infections. The protein

XX can also be used for regulating immune responses, e.g. as antitumor

XX agents, vaccine adjuvants, organ rejection inhibitors or agents for

XX treating autoimmune diseases. The protein can further be used to produce

XX therapeutic antibodies.

AAE77541/C
 ID AAE77541 standard; DNA; 2585 BP.
 XX AAE77541;
 AC AAE77541;
 XX 23-MAY-2001 (first entry)
 DT 23-MAY-2001 (first entry)
 XX Caenorhabditis briggsae mab-21 coding sequence.
 DE Caenorhabditis briggsae mab-21 coding sequence.
 XX Mab-21; cell fate choice; pax-6; aniridia; Moebius syndrome;
 KW chromosome 13q13; peripheral neural tissue differentiation; ds.
 XX Caenorhabditis briggsae.
 XX US6197504-B1.
 PN 06-MAR-2001.
 XX 19-JAN-1998; 98US-0008697.
 PF 10-APR-1996; 96US-0011607.
 PR 09-APR-1997; 97US-0835604.
 XX (CHOW/) CHOW K L.
 PA Chow KL;
 PI Chow KL;
 XX WPI: 2001-256362/26.
 DR P-PSDB; AAB74419.
 XX Detecting expression of mab-21 gene of *Caenorhabditis elegans* encoding
 PT novel protein required for choice of alternate cell fates in sample, by
 PT contacting mRNA from sample with nucleic acid hybridizing with mab-21 -
 XX disclosure: column 41-44; 79pp; English.
 PS The present invention describes a method of detecting mab-21 expression
 XX in a sample involving contacting total mRNA from the sample with a probe
 CC to the mab-21 gene. The sequences of the mab-21 coding sequence and
 CC protein from several species are also given. The human mab-21 gene is
 CC found on chromosome 13q13, a region which has been linked to Moebius
 CC syndrome. The mab-21 protein shows homology to pax-6, mutations in which
 CC are associated with aniridia and which is linked to peripheral neural
 CC tissue differentiation.
 CC Sequence 2585 BP; 751 A; 540 C; 559 G; 688 T; 47 other;

Query Match 56.0%; Score 14; DB 22; Length 2585;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaatt 21
 |||||

Db 1058 CAAGTCACAGCAATP 1045

RESULT 38

AAC69462/C

ID AAC69462 standard; cDNA; 2845 BP.

XX AAC69462;

XX 30-JAN-2001 (first entry)

XX Human secreted protein gene 8 SEQ ID NO:18.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritis;

XX antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;

XX cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;

XX fungicide; ophthalmological; gene therapy; autoimmune disease; infection;

XX hyperproliferative disorder; cardiovascular disorder; angiodysgenesis;

KW cerebrovascular disorder; nervous system disorder; ocular disorder;

PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RR;
XX
XX WPI: 2001-442253/47.
DR P-PSDB: AAM41933.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PS
PS Claim 1: SEQ ID NO 5078; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 5662 BP; 1631 A; 1126 C; 1247 G; 1658 T; 0 other;
SQ

Query Match 56.0%; Score 14; DB 22; Length 5662;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaat 20
| | | | | | | | | |
DB 4497 GCAAGTCACAGAAT 4484

RESULT 41
AAC59953
ID AAC59953 standard; cDNA: 5684 BP.
XX
XX AAC59953:
XX
XX 30-JAN-2001 (first entry)
XX
XX Human secreted protein cDNA sequence #47.
XX
XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
XX antitumor; hepatotropic; antidiabetic; antiinflammatory; antilucer;
XX antiviral; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted; cardiovascular ss.
XX
XX Homo sapiens.
XX
XX WO200055198-A1.
XX
XX 21-SEP-2000.
XX
XX

PF 09-MAR-2000; 2000MO-US06012.
XX
XX 12-MAR-1999; 99US-0124093.
PR 23-NOV-1999; 99US-0166989.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
PI
XX
XX WPI: 2000-587520/55.
DR P-PSDB: AAB28748.
XX
XX Human secreted proteins and the nucleic acids that encode them, useful
PT in gene therapy protocols and recombinant nucleic acid based procedures
PT
PS
PS Claim 1: Page 343-344; 391pp; English.
XX
XX The invention relate to the isolation of genes AAC59907-C59956 encoding
CC 50 human secreted proteins AAB28702-B28751. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
XX Sequence 5684 BP; 1641 A; 1250 C; 1045 G; 1732 T; 16 other;
SQ

Query Match 56.0%; Score 14; DB 21; Length 5684;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagattattt 25
| | | | | | | | | |
DB 4933 tcacagattattt 4946

RESULT 42
AAC76014/C
ID AAC76014 standard; cDNA: 5773 BP.
XX
XX AAC76014:
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
XX
XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX antiviral; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiact;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX antihaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disease; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX

KM thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB41805.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 2354-2357; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
 CC antipsoriatic; antiparinsonian; nootropic; neuroprotective;
 CC osteoprotic; anticonvulsant; antiallergic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; antifungal; antineumatic;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antithrombotic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hyperension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;
 XX
 Query Match 56.0%; Score 14; DB 21; Length 5773;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 OY 7 gcaagtcacagaat 20
 DB 4611 GCAAGTCACAGAAT 4598
 XX
 RESULT 43
 AA159303/C
 ID AA159303 standard; cDNA; 5876 BP.
 XX
 AC AA159303;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1506.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KM Leukemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM40147.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 1506; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activating/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, Leukemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 5876 BP; 1704 A; 1165 C; 1288 G; 1719 T; 0 other;
 XX
 Query Match 56.0%; Score 14; DB 22; Length 5876;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 OY 7 gcaagtcacagaat 20
 DB 4739 GCAAGTCACAGAAT 4726
 XX
 RESULT 44
 AAH24651/C
 ID AAH24651 standard; cDNA; 7063 BP.
 XX
 AC AAH24651;
 XX
 DT 06-AUG-2001 (first entry)
 XX


```

XX XX Nucleotide sequence of a disrupted in schizophrenia 1 (DIS1) gene.
DE XX
XX Disrupted in schizophrenia 1 gene: DIS1 gene; schizophrenia:
KM XX psychiatric disorder: ss.
XX XX
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH 54..2618
FT CDS /*tag= a
FT /*product= "disrupted in schizophrenia 1 (DIS1) protein"
XX
XX MO200140301-A2.
XX
XX 07-JUN-2001.
XX
XX 28-NOV-2000; 2000WO-EP11915.
XX
XX 01-DEC-1999; 99EP-0309667.
XX
XX (ALKU ) AKZO NOBEL NV.
PA (MEDI-) MEDICAL RES COUNCIL.
XX (UYED-) UNIV EDINBURGH.
XX
XX Porteous D, Millar K, Blackwood D;
XX
XX WPI: 2001-374796/39.
DR P-PSDB; AAB84217.
XX
XX Novel isolated polynucleotide which surrounds a breakpoint on
PT chromosome 1 involved in a balanced t(1:11)(q42.1;q14.3)
PT translocation, and its encoded proteins, useful as medication for
PT treating psychiatric disorders
XX
XX Claim 2: Page 30-32; 51pp; English.
XX
XX The present sequence encodes a human disrupted in schizophrenia 1 (DIS1)
CC protein. DIS1 is disrupted by a (1:11)(q42.1;q14.3) translocation which
CC segregates with schizophrenia. The DIS1 gene consists of 13 exons. Exon
CC 11 contains an alternative splice site which removed 66 nucleotides from
CC the open reading frame. The final intron of DIS1 belongs to the rare
CC AT-AC class of introns. DIS1 polypeptide fragments and antibodies are
CC useful for in vitro diagnosis of a psychiatric disorder. DIS1 protein is
CC useful in a screening assay for identifying new drugs for treating
CC psychiatric disorders. DIS1 polypeptides and polynucleotides are useful
CC for preventing and/or treating diseases related to schizophrenia.
CC
XX Sequence 7063 BP; 1930 A; 1644 C; 1690 G; 1799 T; 0 other;
SQ
XX
XX Query Match 56.0%; Score 14; DB 22; Length 7063;
XX Best Local Similarity 100.0%; Pred. No. 35;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 12 tcacagaattattt 25
XX |||||||||||||||
DB 5149 TCACAGAAATTAATT 5136
XX
XX RESULT 45
XX AAT10105
XX ID AAT10105 standard; DNA: 8041 BP.
XX AC
XX AAT10105;
XX
XX 13-MAY-1996 (first entry)
XX
XX Adherence conferring plasmid pear.
XX
XX adhesin; plasmid pear; vector; vaccine; intestine colonisation; ds.
XX
XX Chimeric Escherichia coli;

```

```

OS Chimeric synthetic.
XX
XX Key Location/Qualifiers
FH primer_bind complement (2867..2890)
FT /*tag= a
FT /*note= "primer for E.coli O157:H7 DNA"
XX
XX primer_bind complement (2895..2914)
FT /*tag= b
FT /*note= "primer for pear"
XX
XX CDS 3036..5126
XX /*tag= c
XX /*product= adhesin
XX /*trans_except_pos:3912..3914:..aa:Gly
XX /*note= "claim 1, page 36"
XX
XX misc_feature 3271..3310
XX /*tag= d
XX /*note= "region of transposon TnpHpa insertion"
XX
XX misc_feature 3801..3840
XX /*tag= e
XX /*note= "region of transposon TnpHpa insertion"
XX
XX primer_bind complement (5176..5196)
XX /*tag= f
XX /*note= "primer for pear"
XX
XX primer_bind complement (5159..5183)
XX /*tag= g
XX /*note= "primer for E.coli O157:H7 DNA"
XX
XX CDS complement (6449..7024)
XX /*tag= h
XX /*note= "terf homologue"
XX
XX CDS complement (7092..7670)
XX /*tag= i
XX /*note= "terd homologue"
XX
XX W09600233-A1.
XX
XX 04-JAN-1996.
XX
XX 07-JUN-1995; 95MO-US06994.
XX
XX 24-JUN-1994; 94US-0265714.
XX
XX (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
XX (UNIW ) UNIV WASHINGTON.
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX
XX Besser TE, Bilge SS, Tarr PI, Vary JC;
XX
XX WPI: 1996-068826/07.
DR P-PSDB; AAR75366.
XX
XX Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
PT isolated on plasmid pSC (overlap), for use as a vaccine to prevent
PT bacterial colonisation of bovine intestine
XX
XX Disclosure: Page 23-30; 42pp; English.
XX
XX Adherence conferring plasmid pear (AAT10105) comprises Escherichia
CC coli O157:H7 chromosomal DNA plus the Stratagene SK+ vector.
CC It includes 3 open reading frames, 2 of which are homologues of
CC terE and terF genes necessary for tetracycline resistance. The
CC third ORF is homologous to the 17ga gene and encodes an adhesin
CC (AAR75366) that enables E. coli O157:H7, an antibiotic-resistant,
CC virulent and common food-borne pathogen, to adhere to epithelial
CC cells. This ORF can be utilised in the prodn. of adhesin for use
CC as a vaccine to prevent disease or colonisation of mucosal surfaces
CC by O157:H7.
XX
XX Sequence 8041 BP; 2241 A; 1807 C; 1838 G; 2126 T; 29 other;
SQ
XX
XX Query Match 56.0%; Score 14; DB 17; Length 8041;
XX Best Local Similarity 100.0%; Pred. No. 35;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Thu Dec 27 16:11:57 2001

us-09-396-196f-2.olig.rng

Page 24

Qy 10 agtcacagaattat 23
|||||
Db 3727 agtcacagaattat 3740

Search completed: December 26, 2001, 14:16:32
Job time: 6984 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:52:29 ; Search time 143.1 Seconds
(without alignments)
39,566 Million cell updates/sec

Title: US-09-396-196f-2

Perfect score: 25

Sequence: 1 atgttcgaagtcacagaattatc 25

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 9633

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 100 summaries

Database :

Issued Patents, NA: *
1: /cgn2.6/plodata/2/ina/5A_COMB.seq: *
2: /cgn2.6/plodata/2/ina/6A_COMB.seq: *
3: /cgn2.6/plodata/2/ina/6A_COMB.seq: *
4: /cgn2.6/plodata/2/ina/6B_COMB.seq: *
5: /cgn2.6/plodata/2/ina/6B_COMB.seq: *
6: /cgn2.6/plodata/2/ina/Backlist1.seq: *

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
5	17	68.0	117	2	US-07-944-255-2	Sequence 2, Appli
6	15	60.0	1771	2	US-09-166-203-48	Sequence 48, Appli
7	15	60.0	1771	4	US-09-377-309-48	Sequence 48, Appli
8	14	56.0	1050	6	5498529-7	Patent No. 5498529
9	14	56.0	1677	2	US-08-372-652-14	Sequence 14, Appli
10	14	56.0	1677	5	PCT-US95-16311-14	Sequence 14, Appli
11	14	56.0	1787	2	US-08-372-652-6	Sequence 6, Appli
12	14	56.0	1787	5	PCT-US95-16311-6	Sequence 6, Appli
13	14	56.0	2091	1	US-08-765-081-4	Sequence 4, Appli
14	14	56.0	2091	5	US-09-098-082-4	Sequence 4, Appli
15	14	56.0	2091	5	PCT-US95-06994-4	Sequence 4, Appli
16	14	56.0	2233	1	US-08-496-631-1	Sequence 1, Appli
17	14	56.0	2285	4	US-09-008-697A-7	Sequence 7, Appli
18	14	56.0	8041	4	US-08-765-081-1	Sequence 1, Appli
19	14	56.0	8041	5	US-09-098-082-1	Sequence 1, Appli
20	14	56.0	8041	5	PCT-US95-06994-1	Sequence 1, Appli
21	13	52.0	626	4	US-09-385-982-372	Sequence 372, App
22	13	52.0	626	4	US-09-385-982-250	Sequence 250, App
23	13	52.0	634	4	US-09-385-982-253	Sequence 253, App
24	13	52.0	651	4	US-09-385-982-329	Sequence 329, App
25	13	52.0	1500	4	US-08-875-847B-3	Sequence 3, Appli
26	13	52.0	1505	4	US-08-875-847B-1	Sequence 1, Appli
27	13	52.0	1662	1	US-08-651-572-1	Sequence 1, Appli

28	C	28	13	52.0	1662	3	US-09-066-544-1	Sequence 1, Appli
29	C	29	13	52.0	1662	3	US-08-951-086-1	Sequence 1, Appli
30	C	30	13	52.0	2214	2	US-08-484-993B-1	Sequence 1, Appli
31	C	31	13	52.0	2214	2	US-08-484-158B-1	Sequence 1, Appli
32	C	32	13	52.0	2214	2	US-08-484-596A-1	Sequence 1, Appli
33	C	33	13	52.0	2214	2	US-08-480-150A-1	Sequence 1, Appli
34	C	34	13	52.0	2214	3	US-08-458-731-1	Sequence 1, Appli
35	C	35	13	52.0	2214	3	US-08-149-223A-1	Sequence 9, Appli
36	C	36	13	52.0	2381	2	US-08-484-993B-9	Sequence 9, Appli
37	C	37	13	52.0	2381	2	US-08-484-158B-9	Sequence 9, Appli
38	C	38	13	52.0	2381	2	US-08-484-596A-9	Sequence 9, Appli
39	C	39	13	52.0	2381	2	US-08-480-150A-9	Sequence 9, Appli
40	C	40	13	52.0	2381	3	US-08-458-731-9	Sequence 9, Appli
41	C	41	13	52.0	2381	3	US-08-149-223A-9	Sequence 9, Appli
42	C	42	13	52.0	3083	1	US-07-718-575-5	Sequence 5, Appli
43	C	43	13	52.0	3083	1	US-08-481-206-5	Sequence 5, Appli
44	C	44	13	52.0	3083	2	US-08-486-269A-5	Sequence 5, Appli
45	C	45	13	52.0	4500	2	US-08-743-637B-35	Sequence 35, Appli
46	C	46	13	52.0	4500	2	US-08-526-840B-35	Sequence 35, Appli
47	C	47	13	52.0	5590	4	US-09-050-159-129	Sequence 129, App
48	C	48	13	52.0	72928	3	US-09-009-913-1	Sequence 1, Appli
49	C	49	13	52.0	80266	4	US-09-078-294-4	Sequence 4, Appli
50	C	50	13	52.0	80395	4	US-09-078-294-3	Sequence 3, Appli
51	C	51	12	48.0	17	1	US-08-373-124A-284	Sequence 284, App
52	C	52	12	48.0	17	1	US-08-373-124A-286	Sequence 286, App
53	C	53	12	48.0	17	1	US-08-373-124A-2333	Sequence 2333, Ap
54	C	54	12	48.0	17	1	US-08-373-124A-2335	Sequence 2335, Ap
55	C	55	12	48.0	17	1	US-08-435-628-284	Sequence 284, App
56	C	56	12	48.0	17	1	US-08-435-628-286	Sequence 286, App
57	C	57	12	48.0	17	1	US-08-435-628-2333	Sequence 2333, Ap
58	C	58	12	48.0	17	1	US-08-435-628-2335	Sequence 2335, Ap
59	C	59	12	48.0	24	3	US-09-014-065-15	Sequence 15, Appl
60	C	60	12	48.0	27	1	US-08-373-124A-16	Sequence 16, Appl
61	C	61	12	48.0	27	1	US-08-435-628-16	Sequence 16, Appl
62	C	62	12	48.0	99	1	US-07-795-859B-22	Sequence 22, Appl
63	C	63	12	48.0	99	1	US-08-457-616-22	Sequence 22, Appl
64	C	64	12	48.0	184	2	US-08-485-657A-8	Sequence 8, Appli
65	C	65	12	48.0	184	5	PCT-US95-02303-8	Sequence 8, Appli
66	C	66	12	48.0	255	5	PCT-US93-03077-7	Sequence 7, Appli
67	C	67	12	48.0	279	4	US-08-047-041A-17	Sequence 17, Appl
68	C	68	12	48.0	308	4	US-09-232-522-3	Sequence 3, Appli
69	C	69	12	48.0	311	1	US-08-047-041A-4	Sequence 4, Appli
70	C	70	12	48.0	311	2	US-08-795-006A-24	Sequence 24, Appl
71	C	71	12	48.0	311	4	US-09-184-073-24	Sequence 24, Appl
72	C	72	12	48.0	318	1	US-08-248-874-51	Sequence 51, Appl
73	C	73	12	48.0	318	3	US-08-756-849-51	Sequence 51, Appl
74	C	74	12	48.0	354	3	US-09-065-059-12	Sequence 12, Appl
75	C	75	12	48.0	357	3	US-08-767-128-17	Sequence 17, Appl
76	C	76	12	48.0	411	1	US-08-137-117B-30	Sequence 30, Appl
77	C	77	12	48.0	411	1	US-08-436-717-30	Sequence 30, Appl
78	C	78	12	48.0	420	4	US-09-214-095D-109	Sequence 109, App
79	C	79	12	48.0	426	3	US-08-554-840-28	Sequence 28, Appl
80	C	80	12	48.0	427	2	US-08-484-956-111	Sequence 111, App
81	C	81	12	48.0	427	2	US-08-757-653-111	Sequence 111, App
82	C	82	12	48.0	498	2	US-08-484-956-113	Sequence 113, App
83	C	83	12	48.0	498	2	US-08-757-653-113	Sequence 113, App
84	C	84	12	48.0	500	4	US-08-642-274D-18	Sequence 18, Appl
85	C	85	12	48.0	500	4	US-08-952-014C-18	Sequence 18, Appl
86	C	86	12	48.0	519	5	US-08-581-528A-5	Sequence 5, Appli
87	C	87	12	48.0	519	5	PCT-US94-07799-5	Sequence 5, Appli
88	C	88	12	48.0	600	1	US-08-540-077-1	Sequence 1, Appli
89	C	89	12	48.0	601	2	US-08-484-956-97	Sequence 97, Appl
90	C	90	12	48.0	601	2	US-08-484-956-98	Sequence 98, Appl
91	C	91	12	48.0	601	2	US-08-484-956-99	Sequence 99, Appl
92	C	92	12	48.0	601	2	US-08-484-956-100	Sequence 100, App
93	C	93	12	48.0	601	2	US-08-484-956-101	Sequence 101, App
94	C	94	12	48.0	601	2	US-08-484-956-102	Sequence 102, App
95	C	95	12	48.0	601	2	US-08-484-956-107	Sequence 107, App
96	C	96	12	48.0	601	2	US-08-484-956-108	Sequence 108, App
97	C	97	12	48.0	601	2	US-08-484-956-109	Sequence 109, App
98	C	98	12	48.0	601	2	US-08-484-956-110	Sequence 110, App
99	C	99	12	48.0	601	2	US-08-757-653-97	Sequence 97, Appl
100	C	100	12	48.0	601	2	US-08-757-653-98	Sequence 98, Appl

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgcgaagtcacagaattatt 25
    |||
Db 24 ATTGTCGACAGTCACAGAATTATT 48

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Melgs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEtical: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgcgaagtcacagaattatt 25
    |||
Db 24 ATTGTCGACAGTCACAGAATTATT 48

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biaa"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
OTHER INFORMATION: aminotransf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="biaa RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter plac"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /number=1
DOCUMENT NUMBER: NO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgtcgcaagtcacagaattatt 25
Db 140 attgtcgcaagtcacagaattatt 164

RESULT 4
US-08-411-768B-6
Sequence 6: Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: OF Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
Version 5.1

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,7688
FILING DATE: 31-March-95
CLASSIFICATION: 4.35
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: PB030A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTP synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biOD"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biOD RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6
Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 atgtcgcaagtcacagaattattt 25
|||||
Db 140 attgtcgcaagtcacagaattattt 164

RESULT 5
US-07-944-259-2/c
; Sequence 2, Application US/07944259
; Patent No. 5885792

GENERAL INFORMATION:
APPLICANT: Ifuku, Ohji
APPLICANT: Haze, Shunjiro
APPLICANT: Kishimoto, Jiro
APPLICANT: Nakahama, Kazuo
TITLE OF INVENTION: BIOTIN OPERON
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wegner, Cantor, Mueller & Player
STREET: 123 20th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-8218
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/944,259
FILING DATE: 19920914
CLASSIFICATION: 4.35
ATTORNEY/AGENT INFORMATION:
NAME: Cantor, Herbert I.
REGISTRATION NUMBER: 24,392
REFERENCE/DOCKET NUMBER: P-450-23557
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-0400
TELEFAX: 202-835-0605
TELEX: 440706
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-07-944-259-2

Query Match 68.0%; Score 17; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gtgcgaagtcacagaat 20
|||||
Db 17 gtgcgaagtcacagaat 1

RESULT 6
US-09-166-203-48
; Sequence 48, Application US/09166203A
; Patent No. 5968826
GENERAL INFORMATION:
APPLICANT: Bennett, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowser, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0123
CURRENT APPLICATION NUMBER: US/09/166,203A
CURRENT FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 60
SEQ ID NO 48
LENGTH: 1771
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1193)..(1387)
FEATURE:

NAME/KEY: CDS
LOCATION: (1709)..(1771)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: L20788 Genbank
DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48

Query Match 60.0%; Score 15; DB 2; Length 1771;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 gcaagtcacagaatt 21
|||||
Db 312 gcaagtcacagaatt 326

RESULT 7
US-09-377-309-48
Sequence 48, Application US/09377309B
Patent No. 6258790
GENERAL INFORMATION:
APPLICANT: Bennett, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Coswert, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0390
CURRENT APPLICATION NUMBER: US/09/377,309B
CURRENT FILING DATE: 1999-08-19
EARLIER APPLICATION NUMBER: 09/166,203
EARLIER FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 99
SEQ ID NO 48
LENGTH: 1771
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1193)..(1387)
FEATURE:
NAME/KEY: CDS
LOCATION: (1709)..(1771)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: L20788 Genbank
DATABASE ENTRY DATE: 1996-04-18
US-09-377-309-48

Query Match 60.0%; Score 15; DB 4; Length 1771;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 gcaagtcacagaatt 21
|||||
Db 312 gcaagtcacagaatt 326

RESULT 8
5498529-7/c
Patent No. 5498529
APPLICANT: BERKA, THOMAS R.; FORMALD, JAMES A.; CORNIAK,
JOSEFINA G.; ROSENBERG, MARTIN; STRICKLER, JAMES E.; TAYLOR, DEAN P.
TITLE OF INVENTION: PROTEIN PROTEASE INHIBITORS FROM
STREPTOMYCES
NUMBER OF SEQUENCES: 9
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,506
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 894,167
FILING DATE: 03-JUN-1992
APPLICATION NUMBER: 346,119
FILING DATE: 16-FEB-1989

APPLICATION NUMBER: 897,245
FILING DATE: 18-AUG-1986
SEQ ID NO: 7
LENGTH: 1050
5498529-7

Query Match 56.0%; Score 14; DB 6; Length 1050;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 aagtcacagaatt 22
|||||
Db 251 AAGTCACAGATT 238

RESULT 9
US-08-372-652-14/c
Sequence 14, Application US/08372652
Patent No. 5932699
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Choi, Hwang-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1677 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-372-652-14

Query Match 56.0%; Score 14; DB 2; Length 1677;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ttgtcgcaagtcac 15
|||||
Db 630 TTGTCGCAAGTCAC 617

RESULT 10
PCT-US95-16311-14/c
Sequence 14, Application PC/TUS9516311
GENERAL INFORMATION:

APPLICANT: Moore, David
APPLICANT: Seol, Wongi
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16311
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1677 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-16311-14

Query Match 56.0%; Score 14; DB 5; Length 1677;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ttgtcgcaagtcac 15
|||||

Db 630 TTGTCCCAAGTCAC 617

RESULT 11
US-08-372-652-6/c
Sequence 6, Application US/08372652
Patent No. 5932699
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wongi
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1787 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-372-652-6

Query Match 56.0%; Score 14; DB 2; Length 1787;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ttgtcgcaagtcac 15
|||||

Db 740 TTGTCCCAAGTCAC 727

RESULT 12
PCT-US95-16311-6/c
Sequence 6, Application PC/TUS9516311
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wongi
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16311
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1787 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA
PCT-US95-16311-6

Query Match
Best Local Similarity 100.0%; Score 14; DB 5; Length 1787;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtcgaagtcac 15
|||||
DB 740 TTGTCGACAGTCAC 727

RESULT 13
US-08-765-081-4
Sequence 4, Application US/08765081
Patent No. 5798260

GENERAL INFORMATION:
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.

TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen, O'Connor, Johnson and Kindness

STREET: 2800 Pacific First Center, 1420 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage

COMPUTER: IBM PC compatible/Pentium

OPERATING SYSTEM: MS-Windows 3.1

SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/765,081

FILING DATE: March 26, 1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/06994

FILING DATE: June 7, 1995

APPLICATION NUMBER: US 08/265,714

FILING DATE: June 24, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Shelton, Dennis K.

REGISTRATION NUMBER: 26,997

REFERENCE/DOCKET NUMBER: CHOR-1-10286

TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)

TELEFAX: 1-206-224-0779

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2091 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

DESCRIPTION: Corresponds to SEQ ID NO:1,

DESCRIPTION: nucleotides 3036-5126

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia coli 0157: H7

STRAIN: 86-24 NALR

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2088

US-08-765-081-4

DB 692 AGTCACAGATTAT 705
|||||

RESULT 14
US-09-098-082-4
Sequence 4, Application US/09098082
Patent No. 604042;

GENERAL INFORMATION:
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.

TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen, O'Connor, Johnson and Kindness

STREET: 2800 Pacific First Center, 1420 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage

COMPUTER: IBM PC compatible/Pentium II

OPERATING SYSTEM: MS-Windows 95

SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/098,082

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/765,081

FILING DATE: March 26, 1997

APPLICATION NUMBER: PCT/US95/06994

FILING DATE: June 7, 1995

APPLICATION NUMBER: US 08/265,714

FILING DATE: June 24, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Sheiness, Diana K.

REGISTRATION NUMBER: 35,356

REFERENCE/DOCKET NUMBER: CHOR-1-12402

TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)

TELEFAX: 1-206-224-0779

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2091 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

DESCRIPTION: Corresponds to SEQ ID NO:1,

DESCRIPTION: nucleotides 3036-5126

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia coli 0157: H7

STRAIN: 86-24 NALR

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2088

US-09-098-082-4

Query Match
Best Local Similarity 100.0%; Score 14; DB 3; Length 2091;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattat 23
|||||
DB 692 AGTCACAGATTAT 705

RESULT 15
PCT-US95-06994-4
Sequence 4, Application PC/TUS9506994

GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
ADDRESS: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, Phillip I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINNESS PLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: corresponds to SEQ ID NO:1,
HYPOHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157: H7
STRAIN: 86-24 NALR
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2088
PCT-US95-06994-4

Query Match 56.0%; Score 14; DB 5; Length 2091;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 agtcacagaattat 23
|||||
Db 692 AGTCACAGATTAT 705

RESULT 16
US-08-496-631-1/c
Sequence 1, Application US/08496631
Patent No. 5728548
GENERAL INFORMATION:
APPLICANT: Bowman, Michael
TITLE OF INVENTION: STEROID RECEPTOR RRI
NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:
ADDRESSEE: Genelics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,631
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15248
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2233 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 363..1778
US-08-496-631-1

Query Match 56.0%; Score 14; DB 1; Length 2233;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ttgtcgcaagtcac 15
|||||
Db 1055 TTGTGCGCAAGTCAC 1042

RESULT 17
US-09-008-697A-7/c
Sequence 7, Application US/09008697A
Patent No. 6197504
GENERAL INFORMATION:
APPLICANT: Chow, King Lau
TITLE OF INVENTION: USES OF MAB-21
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,697A
FILING DATE: January 19, 1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chan, Albert Wai-Kit
REGISTRATION NUMBER: 36,479

```
REFERENCE/DOCKET NUMBER: 50752-A
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (212) 278-0400
  TELEFAX: (212) 391-0525
  INFORMATION FOR SEQ ID NO: 7:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 2585 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
US-09-008-697A-7

Query Match
Best Local Similarity 56.0%; Score 14; DB 4; Length 2585;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaatt 21
DB 1058 CAAGTCACAGAATT 1045

RESULT 18
US-08-765-081-1
Sequence 1, Application US/08765081
Patent No. 5798260
GENERAL INFORMATION:
  APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
  TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin
  NUMBER OF SEQUENCES: 12
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
    STREET: 2800 Pacific First Center, 1420 Fifth Avenue
    CITY: Seattle
    STATE: Washington
    COUNTRY: USA
    ZIP: 98101-2347
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
    COMPUTER: IBM PC compatible/Pentium
    OPERATING SYSTEM: MS-Windows 3.1
    SOFTWARE: Word for Windows-6.0
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/765,081
    FILING DATE: March 26, 1997
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/US95/06994
      FILING DATE: June 7, 1995
      APPLICATION NUMBER: US 08/265,714
      FILING DATE: June 24, 1994
      ATTORNEY/AGENT INFORMATION:
        NAME: Shelton, Dennis K.
        REGISTRATION NUMBER: 26,997
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: 1-206-682-8160; 1-206-224-0718 (direct)
        TELEFAX: 1-206-224-0779
        INFORMATION FOR SEQ ID NO: 1:
          SEQUENCE CHARACTERISTICS:
            LENGTH: 8041 base pairs
            TYPE: nucleic acid
            STRANDEDNESS: single
            TOPOLOGY: linear
          MOLECULE TYPE: DNA (genomic)
          HYPOTHETICAL: NO
          ANTI-SENSE: NO
          ORIGINAL SOURCE:
            ORGANISM: Escherichia coli O157:H7
            STRAIN: 86-24 NALR
            IMMEDIATE SOURCE:
              CLONE: PEAR
US-08-765-081-1

Query Match
Best Local Similarity 56.0%; Score 14; DB 1; Length 8041;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattat 23
DB 3727 AGTCACAGAATTAT 3740

RESULT 19
US-09-098-082-1
Sequence 1, Application US/09098082
Patent No. 6040421
GENERAL INFORMATION:
  APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
  TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin
  NUMBER OF SEQUENCES: 12
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
    STREET: 2800 Pacific First Center, 1420 Fifth Avenue
    CITY: Seattle
    STATE: Washington
    COUNTRY: USA
    ZIP: 98101-2347
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
    COMPUTER: IBM PC compatible/Pentium II
    OPERATING SYSTEM: MS-Windows 95
    SOFTWARE: Word for Windows-6.0
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/098,082
    FILING DATE:
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/765,081
        FILING DATE: March 26, 1997
        APPLICATION NUMBER: PCT/US95/06994
        FILING DATE: June 7, 1995
        APPLICATION NUMBER: US 08/265,714
        FILING DATE: June 24, 1994
        ATTORNEY/AGENT INFORMATION:
          NAME: Sheiness, Diana K.
          REGISTRATION NUMBER: 35,356
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
          TELEFAX: 1-206-224-0779
          INFORMATION FOR SEQ ID NO: 1:
            SEQUENCE CHARACTERISTICS:
              LENGTH: 8041 base pairs
              TYPE: nucleic acid
              STRANDEDNESS: single
              TOPOLOGY: linear
            MOLECULE TYPE: DNA (genomic)
            HYPOTHETICAL: NO
            ANTI-SENSE: NO
            ORIGINAL SOURCE:
              ORGANISM: Escherichia coli O157:H7
              STRAIN: 86-24 NALR
              IMMEDIATE SOURCE:
                CLONE: PEAR
US-09-098-082-1

Query Match
Best Local Similarity 56.0%; Score 14; DB 3; Length 8041;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattat 23
DB 3727 AGTCACAGAATTAT 3740
```

RESULT 20
PCT-US95-06994-1
Sequence 1, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, Phillip I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
IMMEDIATE SOURCE:
CLONE: PEAR
PCT-US95-06994-1

Query Match 56.0%; Score 14; DB 5; Length 8041;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 agtcacagaattat 23
|||||
Db 3727 agtcacagaattat 3740

RESULT 21
US-09-385-982-372/c
Sequence 372, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O, ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

TITLE OF INVENTION: PRODUCTS: II
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 372
LENGTH: 607
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(607)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-372

Query Match 52.0%; Score 13; DB 4; Length 607;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagaattat 24
|||||
Db 189 tcacagaattatt 177

RESULT 22
US-09-385-982-250
Sequence 250, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O, ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 08/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 250
LENGTH: 626
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(626)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-250

Query Match 52.0%; Score 13; DB 4; Length 626;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagaattat 24
|||||
Db 177 tcacagaattatt 189

RESULT 23
US-09-385-982-253

Sequence 253, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 253
LENGTH: 634
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(634)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-253

Query Match 52.0%; Score 13; DB 4; Length 634;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagaattatt 24
|||||

Db 196 tcacagaattatt 208

RESULT 24
US-09-385-982-329
Sequence 329, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 329
LENGTH: 651
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(651)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-329

Query Match 52.0%; Score 13; DB 4; Length 651;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 12 tcacagaattatt 24
|||||

Db 169 tcacagaattatt 181

RESULT 25
US-08-875-847B-3/c
Sequence 3, Application US/08875847B
Patent No. 6255105
GENERAL INFORMATION:
APPLICANT: The Government of the United States of America as represented by the Secretary, Department of Health and Human Services; Callahan, Robert; Marchetti, Applicant; Antonio; Buttila, Fiamma; Smith, Gilbert H.
TITLE OF INVENTION: Nucleotide And Deduced Amino Acid Sequences Of A New Tumor Gene, Int6, And the Use Of Reagents Derived From These Sequences In Diagnostic Assays, Vaccines, Immunotherapy And Gene Therapy
TITLE OF INVENTION: Int6, And the Use Of Reagents Derived From These Sequences In Diagnostic Assays, Vaccines, Immunotherapy And Gene Therapy
TITLE OF INVENTION: Int6, And the Use Of Reagents Derived From These Sequences In Diagnostic Assays, Vaccines, Immunotherapy And Gene Therapy
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/875,847B
FILING DATE: 09-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/385,998
FILING DATE: 09-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: William S. Feller
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4179PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-875-847B-3

Query Match 52.0%; Score 13; DB 4; Length 1500;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagaattatt 24
|||||

Db 624 TCACAGATTATT 612

RESULT 26
US-08-875-847B-1/c
Sequence 1, Application US/08875847B
Patent No. 6255105
GENERAL INFORMATION:
APPLICANT: The Government of the United States of America as represented by the Secretary, Department of Health and Human

APPLICANT: Services; Callahan, Robert; Marchetti,
APPLICANT: Antonio; Buttila, Flamma; Smith, Gilbert H.
TITLE OF INVENTION: Nucleotide And Deduced
TITLE OF INVENTION: Amino Acid Sequences Of A New Tumor Gene,
TITLE OF INVENTION: Int6, And The Use Of Reagents Derived From
TITLE OF INVENTION: These Sequences In Diagnostic Assays,
TITLE OF INVENTION: Vaccines, Immunotherapy And Gene Therapy
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/875,847B
FILING DATE: 09-FEB-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/385,998
FILING DATE: 09-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: William S. Feller
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4179PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1505 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-875-847B-1

Query Match 52.0%; Score 13; DB 4; Length 1505;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tcacagaattatt 24
|||||
Db 630 TCACAGAAATTATT 618

RESULT 27
US-08-651-572-1/c
Sequence 1, Application US/08651572
Patent No. 5789228
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,572
FILING DATE: Herewith
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-08-651-572-1

Query Match 52.0%; Score 13; DB 1; Length 1662;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaatt 21
|||||
Db 974 AAGTCACGAATT 962

RESULT 28
US-09-066-544-1/c
Sequence 1, Application US/09066544
Patent No. 6001984
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,544
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/651,572
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES

TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-09-066-544-1

Query Match 52.0%; Score 13; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 aagtcacagaatt 21
|||||
Db 974 AAGTCACAGATT 962

RESULT 29
US-08-951-086-1/c
Sequence 1, Application US/08951086
Patent No. 6074867
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglycanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CECCHI, STEWART & OLSTEIN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,086
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/651,572
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Hefton
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-08-951-086-1

Query Match 52.0%; Score 13; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 aagtcacagaatt 21
|||||
Db 974 AAGTCACAGATT 962

RESULT 30
US-08-484-993B-1/c
Sequence 1, Application US/08484993B

Patent No. 5837497
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunoreception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,993B
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: oocyte
FEATURE:
NAME/KEY: sia_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-993B-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 gcaatcacaga 19
|||||
Db 1898 GCAATCAGAGA 1886

RESULT 31
US-08-484-158B-1/c
Sequence 1, Application US/08484158B
Patent No. 5976545
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Pharmaceutical Compositions for
NUMBER OF SEQUENCES: 61
TITLE OF INVENTION: Immunoncontraception
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,158B
FILING DATE: 07-JUNE-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 09-NOV-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-92
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 32794
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-158B-1

Query Match 52.0%; Score 13; DB 2; Length 2214;

Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 gcaagtcacagaa 19
Db 1898 GCAAGTCACAGAA 1886

RESULT 32
US-08-484-596A-1/c
Sequence 1, Application US/08484596A
Patent No. 5981228
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunoncontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,596A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 11-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-596A-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 33
US-08-480-150A-1/c
; Sequence 1, Application US/08480150A
; Patent No. 5989550
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunoccontraception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/480,150A
; APPLICATION NUMBER: US/08/480,150A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/149,223
; FILING DATE: 09-NOV-1993
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2214 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Sus scrofa
; DEVELOPMENTAL STAGE: Juvenile
; HAPLOTYPE: Diploid
; TISSUE TYPE: Ovary
; CELL TYPE: Oocyte
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 12..119
; FEATURE:
; NAME/KEY: mat_peptide

; LOCATION: 120..2153
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 12..2153
; US-08-480-150A-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 34
US-08-458-731-1/c
; Sequence 1, Application US/08458731
; Patent No. 6001599
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunoccontraception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/458,731
; APPLICATION NUMBER: US/08/458,731
; FILING DATE: 09-NOV-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2214 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Sus scrofa
; DEVELOPMENTAL STAGE: Juvenile
; HAPLOTYPE: Diploid
; TISSUE TYPE: Ovary
; CELL TYPE: Oocyte
; FEATURE:
; NAME/KEY: sig_peptide

LOCATION: 12..119
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-458-731-1

Query Match 52.0%; Score 13; DB 3; Length 2214;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaa 19
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 35
US-08-149-223A-1/C
Sequence 1, Application US/08149223A
Patent No. 6027727
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149, 223A
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012, 990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973, 341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPOTYPE: Diploidy
TISSUE TYPE: Ovary

CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig-peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-149-223A-1

Query Match 52.0%; Score 13; DB 3; Length 2214;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaa 19
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 36
US-08-484-993B-9/C
Sequence 9, Application US/08484993B
Patent No. 5837497
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484, 993B
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012, 990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973, 341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris

DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploid
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-993B-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 37
US-08-484-158B-9/c
; Sequence 9, Application US/08484158B
; Patent No. 5976545
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Pharmaceutical Compositions for
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,158B
; FILING DATE: 07-JUNE-95
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/149,223
; FILING DATE: 09-NOV-93
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-93
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 32794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2381 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploid
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-158B-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 38
US-08-484-596A-9/c
; Sequence 9, Application US/08484596A
; Patent No. 5981228
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunoreception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,596A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/149,223
; FILING DATE: 11-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2381 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Canis familiaris
; DEVELOPMENTAL STAGE: Juvenile

HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-596A-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 39
US-08-480-150A-9/c
Sequence 9, Application US/08480150A
Patent No. 5989550
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immuncontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480.150A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/149,223
FILING DATE: 09-NOV-1993
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy

TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-480-150A-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 40
US-08-458-731-9/c
Sequence 9, Application US/08458731
Patent No. 6001599
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immuncontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458.731
FILING DATE: 09-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:

NAME/KEY: CDS
LOCATION: 206..2353
US-08-458-731-9

Query Match 52.0%; Score 13; DB 3; Length 2381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
DB 2098 GCAAGTCACAGAA 2086

RESULT 41
US-08-149-223A-9/c
Sequence 9, Application US/08149223A
Patent No. 6027727
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunoc contraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,223A
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-149-223A-9

Query Match 52.0%; Score 13; DB 3; Length 2381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
DB 2098 GCAAGTCACAGAA 2086

RESULT 42
US-07-718-575-5/c
Sequence 5, Application US/07718575
Patent No. 5202257
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael MN
APPLICANT: Bettler Ph.D., Bernhard MN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Bruggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/718,575
FILING DATE: 19910813
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: GluR3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-07-718-575-5

Query Match 52.0%; Score 13; DB 1; Length 3083;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgcaagtc 13
|||||
DB 3033 ATTGTGCAAGTC 3021

RESULT 43

US-08-481-206-5/c
Sequence 5, Application US/08481206
Patent No. 5739291
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael MN
APPLICANT: Bettler Ph.D., Bernhard MN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,206
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,767
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 910330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: GLUR3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-08-481-206-5

Query Match 52.0%; Score 13; DB 1; Length 3083;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgcaagtc 13
|||||
Db 3033 ATTGTCGCAAGTC 3021

RESULT 44
US-08-486-269A-5/c
Sequence 5, Application US/08486269A
Patent No. 5945509
GENERAL INFORMATION:
APPLICANT: Heinemann, Stephen F.
APPLICANT: Boulter, James R.
APPLICANT: Hollmann, Michael
APPLICANT: Bettler, Bernhard
APPLICANT: Jensen, Jan E.

TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
TITLE OF INVENTION: AND METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,269A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,767
FILING DATE: 04-FEB-1993
APPLICATION NUMBER: 07/718,575
FILING DATE: 21-JUN-1991
APPLICATION NUMBER: PCT/US90/06153
FILING DATE: 25-OCT-1990
APPLICATION NUMBER: 07/428,116
FILING DATE: 27-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9986
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-677-1409
TELEFAX: 619-677-1465
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: GLUR3
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 167...2830
OTHER INFORMATION:
US-08-486-269A-5

Query Match 52.0%; Score 13; DB 2; Length 3083;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

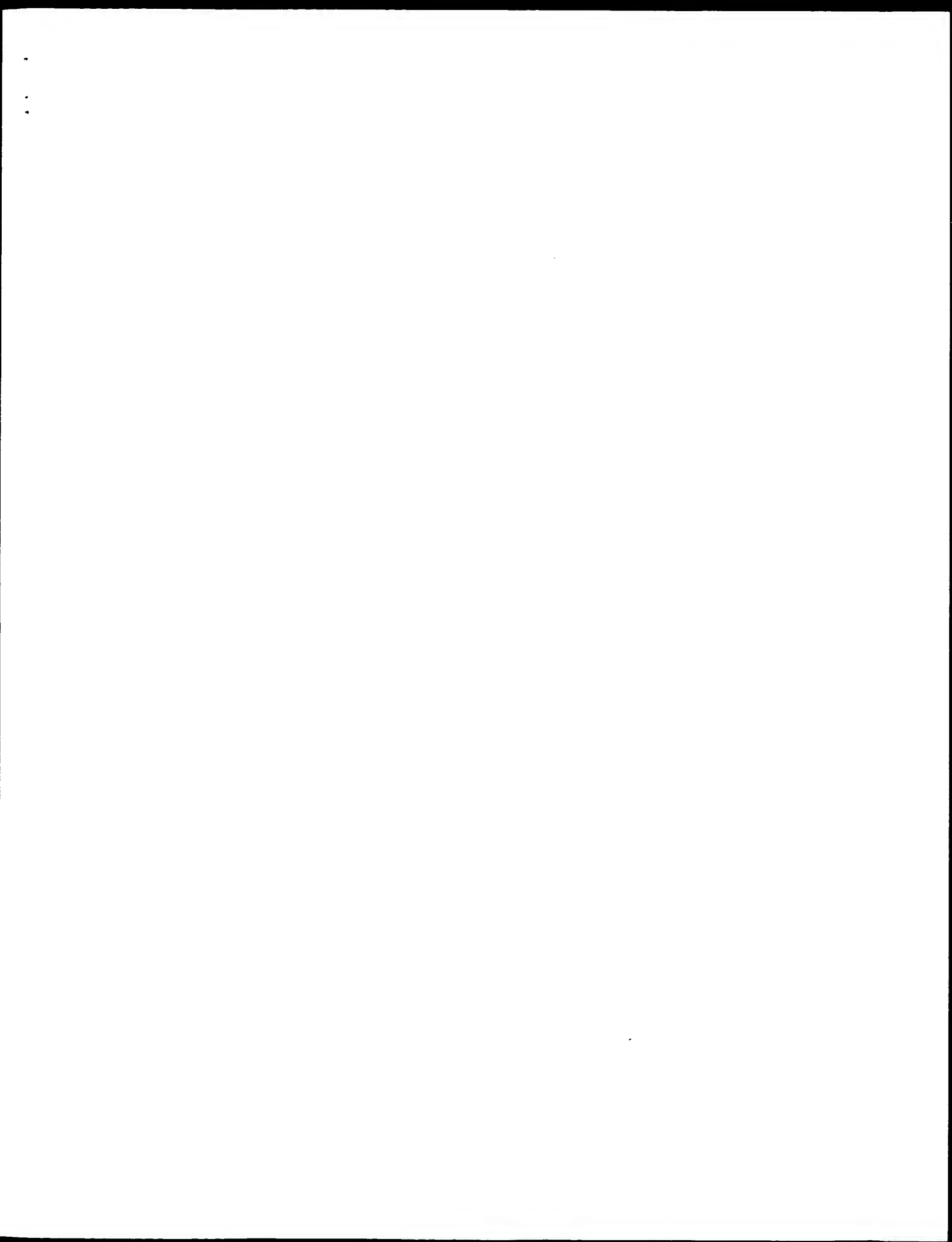
OY 1 attgtcgcaagtc 13
|||||
Db 3033 ATTGTCGCAAGTC 3021

RESULT 45
US-08-743-637B-35/c
Sequence 35, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED

TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: O'ARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ. ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 4500 base pairs
TYPE: nucleic acid
STRADEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
US-08-743-637B-35

Query Match 52.0%; Score 13; DB 2; Length 4500;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9 aagtcacagaatt 21
|||||
Db 1911 AAGTCACAGAATT 1899

Search completed: December 26, 2001, 12:52:31
Job time: 6783 sec




```

c 86 15 60.0 665 13 A2459164
c 87 15 60.0 691 13 A2963904
c 88 15 60.0 782 13 BH028572
c 89 15 60.0 849 13 BE548526
c 90 15 60.0 874 11 BF239747
c 91 15 60.0 912 10 AA626775
c 92 15 60.0 1239 10 BE735021
c 93 14 56.0 114 11 BF987752
c 94 14 56.0 116 10 AA022548
c 95 14 56.0 116 11 BF991166
c 96 14 56.0 132 13 A2073992
c 97 14 56.0 142 10 BB184084
c 98 14 56.0 142 10 BB266371
c 99 14 56.0 149 10 BB300077
c 100 14 56.0 150 10 AW817267

```

ALIGNMENTS

```

RESULT 1
LOCUS BE800706 357 bp mRNA EST 20-SEP-2000
DEFINITION sq97c06.y1 Gm-cl049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl049-1091 5' similar to TR:082150 082150 CHLOROPLAST FTSH
ACCESSION BE800706
VERSION BE800706.1 GI:10231818
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 357)
REFERENCE 1
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
JOURNAL Public Soybean EST Project
COMMENT Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length is 364
High quality sequence stop: 340.
FEATURES
source
Location/Qualifiers
1..357
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl049-1091"
/clone_1lb="Gm-cl049"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site_1: EcoRI, Site_2:
XhoI; The Clark NIL was constructed and seed was provided
by Dr. J. Specht, University of Nebraska (Shoemaker and
Specht, 1995). The cDNA library was constructed from mRNA
isolated from whole seedlings of 3 week old greenhouse

```

```

BASE COUNT 169 a 53 c 47 g 88 t
ORIGIN
gown plants. Complementary DNA was synthesized from mRNA
using a primer consisting of a poly(dT) sequence with a
XhoI restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by XhoI
digestion. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
DH10B host cells (GibcoBRL). The library was constructed
in cooperation with Dr. Paul Keim's laboratory at Northern
Arizona University."

```

```

Query Match 72.0%; Score 18; DB 11; Length 357;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 6 cgcacgtcacagaattat 23
|||||
Db 138 CCGACGTACACGAATTAT 155

```

RESULT 2

```

LOCUS AW781150 388 bp mRNA EST 12-MAY-2000
DEFINITION s189g06.y1 Gm-cl037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl037-1235 5' similar to TR:082150 082150 CHLOROPLAST FTSH
ACCESSION AW781150
VERSION AW781150.1 GI:7795753
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 388)
REFERENCE 1
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
JOURNAL Public Soybean EST Project
COMMENT Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 375.
FEATURES
source
Location/Qualifiers
1..388
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl037-1235"
/clone_1lb="Gm-cl037"
/tissue_type="fully expanded leaves of greenhouse grown
plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="Vector: pSPORT1, Site_1: NotI, Site_2: SalI; This
cDNA library was constructed from mRNA isolated from fully
expanded leaves of greenhouse grown plants that were 2

```

weeks old. The library was prepared using the Lite Technologies psupertscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(OT) sequence with a NotI restriction site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electro-Max DH10B host cells. This library by Ann Khanna at the University of Illinois at Urbana-Champaign. email: l-vodkin@uiuc.edu"

BASE COUNT 123 a 67 c 92 g 106 t
ORIGIN
Query Match 72.0%; Score 18; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ccgaagtcacagaattat 23
Db 334 CCAGAGTCACAGAAATTAT 351

RESULT 3
BE658717/c 751 bp mRNA EST 24-MAY-2001
LOCUS GM700007A10P9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3,
DEFINITION mRNA sequence.
ACCESSION BE658717
VERSION BE658717.1 GI:9984609
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 751)
AUTHORS Vodka, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V., Eppel, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.
TITLE A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL Unpublished (1999)
COMMENT Other ESTs: AW156684 corresponding to Gm-cl015-2632 (5')
Contact: Vodka, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu

FEATURES
SOURCE This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.
Location/Qualifiers
1..751
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1070-2561"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, retracted set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or

a representative of each contig, which were reracked to form library Gm-r1070. The cDNA clones of the reracked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the Laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota. http://www.cbc.umn.edu/research/projects/soybean/index.html . Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT 193 a 171 c 131 g 226 t 30 others
ORIGIN
Query Match 72.0%; Score 18; DB 10; Length 751;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ccgaagtcacagaattat 23
Db 112 CCAGAGTCACAGAAATTAT 95

RESULT 4
AO227679 424 bp DNA GSS 26-SEP-1998
LOCUS HS_2012_B2_F10_T7 CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone plate=2012 Col=20 Row=L, DNA sequence.
ACCESSION AO227679
VERSION AO227679.1 GI:3652908
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 424)
AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2012 row: L column: 20
Class: BAC ends
High quality sequence stop: 424.

FEATURES
SOURCE Location/Qualifiers
1..424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2012 Col=20 Row=L"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 107 a 86 c 87 g 143 t 1 others
ORIGIN

Query Match	64.0%;	Score 16;	DB 10;	Length 299;
Best Local Similarity	100.0%;	Pred. No. 35;		
Matches	16;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;
QY	6	cgcgaagtcacagaatt 21		
Db	3	CGCAAGTCACAGAATT 18		
RESULT	9			
LOCUS	AA648449			
DEFINITION	ns22f10.s1 NCI-CCAP_GCB1 Homo sapiens CDNA clone IMAGE:1184395 3'			
ACCESSION	AA648449			
VERSION	AA648449			
KEYWORDS	AA648449.1 GI:2574878			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 353)			
JOURNAL	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.			
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cga@bbs-rfemail.nih.gov			
	unknown library type			
	Insert Length: 2027 Std Error: 0.00			
	Seq primer: -40m13 fwd. ET from Amersham			
	High quality sequence stop: 334.			
FEATURES	Location/Qualifiers			
source	1..353			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone_id="IMAGE:1184395"			
	/clone_id="NCI-CCAP_GCB1"			
	/tissue_type="germinal center B cell"			
	/lab_host="DH10B"			
	/note="Vector: pT73D-Pac (Pharmacia) with a modified			
	polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA			
	was prepared from human tonsillar cells enriched for			
	germinal center B cells by flow sorting (CD20+, IgD-),			
	provided by Dr. Louis M. Staudt (NCI), Dr. David Allman			
	(NCI) and Dr. Gerald Marli (CBER). cDNA synthesis was			
	primed with a Not I - oligo(dT) primer			
	[5'-GTTCACCAATGTGAAGTGGAGGCGCGCCATTTTGTTTT-TTTT-3'			
] . Double-stranded cDNA was ligated to Eco RI adaptors			
	(Pharmacia) , digested with Not I and cloned into the Not I			
	/tissue_type="spinal ganglion"			
	/dev_stage="12 days embryo"			
	/lab_host="DH10B"			
	/note="Site_1: SalI; Site_2: BamHI; cDNA library was			
	prepared and sequenced in Mouse Genome Encyclopedia			
	Project of Genome Exploration Research Group in Riken			
	Genomic Sciences Center and Genome Science Laboratory in			
	RIKEN. Division of Experimental Animal Research in Riken			
	contributed to prepare mouse tissues. 1st strand cDNA was			
	primed with a primer [5',			
	GAGACAGAGAGCGCGCCGCAACTCGAGTTTGTTTTGTTTTNN 3'], cDNA was			
	prepared by using urealase thermo-activated reverse			
	transcriptase and subsequently enriched for full-length by			
	cap-trapper. Second strand cDNA was prepared with the			
	primer adapter of sequence [5',			
	GAGACAGAGAGTTCTCGACTTAATTAAATATCCCGCCCCCCCC 3']. cDNA			
	was cleaved with BamHI and XhoI. Vector: a modified			
	pBluescript KS(+) after bulk excision from Lambda FLG I."			
BASE COUNT	75	a	72	c
ORIGIN		45	g	107
		t		

BASE COUNT	131 a	49 c	39 g	134 t	
ORIGIN					
Query Match	64.0%	Score 16:	DB 10;	Length 353;	
Best Local Similarity	100.0%	Pred. No. 35;			
Matches 16;	Conservative	0;	Mismatches	0;	Indels 0;
Gaps	0;				
OY	8	caagtcacagaattat 23			
DB	249	CAAGTCACAGAATTAT 264			
RESULT 10					
BF726110					
LOCUS	BF726110	405 bp	mRNA	EST	05-JAN-2001
DEFINITION	by01c01.x1 Human Lens cDNA (Un-normalized, unamplified): By Homo				
ACCESSION	Sapiens cDNA Clone by01c01 3', mRNA sequence.				
VERSION	BF726110				
KEYWORDS	BF726110.1	GI:12042021			
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Eukaryota: Metazoa: Chordata; Craniata; Vertebrata: Euteleostomi:				
AUTHORS	Mammalia; Eutheria: Primates; Catarrhini: Homnidae; Homo.				
TITLE	1 (bases 1 to 405)				
JOURNAL	Wistow,G., Bernstein,S., Behal,A. and Smith,D.				
COMMENT	NETBANK: EST analysis and bioinformatics for ocular genomics				
	Invest. Ophthalmol. Vis. Sci. 41 (2000) In press				
	Contact: Wistow G				
	Section on Molecular Structure and Function				
	National Eye Institute				
	6/331, NIH, Bethesda, MD 20892-2740, USA				
	Tel: 301 402 3452				
	Fax: 301 496 0078				
	Email: graeme@helix.nih.gov				
	Plate: 01 row: c column: 01				
	Seq primer: -21M13 forward primer (ABI).				
FEATURES	Location/Qualifiers				
source	1. 405				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="by01c01"				
	/clone_lib="Human Lens cDNA (Un-normalized, unamplified):				
	By"				
	/rname_type="lens"				
	/dev_stage="adult"				
	/lab_host="EMDH10B"				
	/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses				
	from different adults (both approximately 40 years old)				
	together yielded 20ug of total RNA and 150ng mRNA for cDNA				
	library synthesis. A directionally cloned cDNA library in				
	the pCMVSPORT6 vector was constructed at Life Technologies				
	essentially following the protocols of the Superscript				
	Plasmid System full details of which are contained in the				
	manufacturer's instruction manual				
	(http://www.lifetechn.com/). First strand synthesis was				
	carried out using a Not I primer-adaptor				
	[5'-pGACTAGTCTACATCGCGCGCGCCGCTT15-3']. Not I blunt				
	end inserts were cloned into the Not I/EcoR V sites in the				
	vector. EST analysis was performed on the unamplified				
	library at the NIH Intramural Sequencing Center (NISC)."				
BASE COUNT	140 a	58 c	47 g	160 t	
ORIGIN					
Query Match	64.0%	Score 16:	DB 11;	Length 405;	
Best Local Similarity	100.0%	Pred. No. 35;			
Matches 16;	Conservative	0;	Mismatches	0;	Indels 0;
Gaps	0;				

Plate: 2175 row: B column: 12
Class: BAC ends
High quality sequence stop: 454.

FEATURES

source

Location/Qualifiers

1..454

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Plate=2175 Col=12 Row=B"

/note="Organ: sperm; Vector: pHELOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT

188 a 70 c 76 g 117 t 3 others

ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 454;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaattat 23
|||||

Db 156 CAAGTCACAGAATTAT 171

RESULT 14

A1361386

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

www-bio.llnl.gov/biop/image/image.html

Insert length: 1228 Std Error: 0.00

Seq primer: -400P from Glibco

High quality sequence stop: 434.

Location/Qualifiers

1..457

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:2015161"

/clone_lib="NCI CGAP_Brn23"

/tissue_type="glibolastoma (pooled)"

/lab_host="DH10B"

/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - Oligo(dT) primer [5'

TGTACCAATCATGACGAGCGCGCATCTTTTCTTTTCTTTTCTTTTCTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

Library is normalized, and was constructed by Bento

BASE COUNT 153 a 73 c 52 g 179 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 457;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaattat 23
|||||

Db 153 CAAGTCACAGAATTAT 168

RESULT 15

BG360303

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

www-bio.llnl.gov/biop/image/image.html

Insert length: 1228 Std Error: 0.00

Seq primer: -400P from Glibco

High quality sequence stop: 465.

Location/Qualifiers

1..467

/organism="Xenopus laevis"

/db_xref="taxon:835"

/clone_lib="IMAGE:4405125"

/clone_lib="NICHD XGC Emb2"

/tissue_type="embryo, stage 17/19"

/lab_host="DH10B (phage-resistant)"

/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;

cloned unidirectionally. Primer: Oligo dT. Average insert

size 2.1 kb. Constructed by Life Technologies."

Location/Qualifiers

146 a 86 c 84 g 151 t

BASE COUNT

ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 467;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 agtcacagaattat 25
|||||

Db 89 AGTCACAGAATTATT 104

RESULT 16

A1635259

LOCUS A1635259 471 bp mRNA EST 16-DEC-1999
 DEFINITION tz79h07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2294845 3',
 mRNA sequence.
 ACCESSION A1635259
 VERSION A1635259.1 GI:4686589
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 471)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Life Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrp/image/image.html
 Insert length: 1533 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 399.
 Location/Qualifiers
 1..471
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2294845"
 /clone_lib="NCI_CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1: Salt;
 Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.72 kb. Life Technologies catalog #:
 11548-013"

BASE COUNT 171 a 69 c 53 g 178 t
 ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 471;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacagaattat 23
 |||||||
 Db 235 CAAGTCACAGAAATTAT 250

RESULT 17
 A1634722 503 bp mRNA EST 16-DEC-1999
 LOCUS tz09e04.x1 NCI_CGAP_Utl1 Homo sapiens cDNA clone IMAGE:2288094 3',
 similar to contains element XTR repetitive element.; mRNA
 DEFINITION sequence.
 ACCESSION A1634722 GI:4686052
 VERSION A1634722.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 503)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrp/image/image.html
 Insert length: 3719 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 413.
 Location/Qualifiers
 1..503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2288094"
 /clone_lib="NCI_CGAP_Utl1"
 /tissue_type="well-differentiated endometrial
 adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: Salt;
 Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.75 kb. Life Technologies catalog #:
 11538-014"

BASE COUNT 184 a 77 c 56 g 186 t
 ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 503;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacagaattat 23
 |||||||
 Db 236 CAAGTCACAGAAATTAT 251

RESULT 18
 BG347513 503 bp mRNA EST 28-FEB-2001
 LOCUS dab87h12.x1 NICHD XGC Emb2 Xenopus laevis cDNA clone IMAGE:4404719
 DEFINITION 3', mRNA sequence.
 ACCESSION BG347513
 VERSION BG347513.1 GI:13167937
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodidae; Xenopus.
 1 (bases 1 to 503)
 Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
 Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person,
 B., Gibbons, M., Harvey, N., Riter, E., Jackson, Y., McCann, R.,
 Waterston, R., and Wilson, R.
 WashU Xenopus EST project, 1999
 Unpublished (1999)
 Contact: Sandy Clifton, Ph.D.
 WashU Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library constructed by Life Technologies, DNA Sequencing by:
 Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -400P from Gibco
 High quality sequence stop: 432.
 Location/Qualifiers
 1..503
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:4404719"
 /clone_lib="NICHD XGC Emb2"

FEATURES
 source

```

/lab-host="embryo, stage 17/19"
/clone="Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 kb. Constructed by Life Technologies."
BASE COUNT      167 a      87 c      89 g      160 t
ORIGIN
Query Match      64.0%; Score 16; DB 11; Length 503;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 10 agtcacgaattattt 25
Db 51 AGTCACGAAATTATT 66

```

```

RESULT 19
AM418528      512 bp      mRNA      EST      09-FEB-2000
LOCUS      x132c07.x1 NCI CGAP ut4 Homo sapiens cDNA clone IMAGE:2761836 3'
DEFINITION      similar to contains element MER32 repetitive element ;, mRNA
sequence.
ACCESSION      AM418528      GI:6946410
VERSION      AM418528
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 512)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaabs-rt@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -400p from Gibco
High quality sequence stop: 404.
FEATURES
Source      1..512
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2761836"
/clone_lib="NCI-CGAP-ut4"
/tissue-type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"
BASE COUNT      186 a      79 c      56 g      191 t
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 10; Length 512;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 8 caagtcacgaattat 23
Db 240 CAAgTCACGAAATTAT 255

```

```

RESULT 20
AO472309      527 bp      DNA      GSS      23-APR-1999
LOCUS      CITBI-EL-2587F18.TF CITBI-EL Homo sapiens genomic clone 2587F18,
DEFINITION      DNA sequence.
ACCESSION      AO472309      GI:4655963
VERSION      AO472309.1
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 527)
AUTHORS      Zhao, S., Adams, M.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and
Venter, J.C.
TITLE      Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
Map Building
JOURNAL      Unpublished (1997)
COMMENT      Other_GSSs: CITBI-EL-2587F18.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetlgr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
Source      1..527
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2587F18"
/clone_lib="CITBI-EL"
/sex="male"
/cell-type="sperm"
/note="Vector: pBelosBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT      150 a      88 c      83 g      205 t      1 others
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 13; Length 527;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 10 agtcacgaattattt 25
Db 81 AGTCACGAAATTATT 96

```

```

RESULT 21
BF322232      649 bp      mRNA      EST      21-NOV-2000
LOCUS      maa34h05.y1 NCI-CGAP L110 Mus musculus cDNA clone IMAGE:3812985 5'
DEFINITION      similar to SW:CPBL1MOUSE 005421 CTTCCHROME P450 2E1 ;, mRNA
sequence.
ACCESSION      BF322232      GI:11271563
VERSION      BF322232
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 649)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)

```

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.femail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov

MGI:1455097
Seq primer: -40RP from GIBCO
High quality sequence stop: 434.
Location/Qualifiers
1. 649

FEATURES
source
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3812985"
/clone_lib="NCL_CGAP_L110"
/sex="female"
/dev_stage="10 weeks"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Liver; Vector: pCMV-Sport6; Site_1: NciI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Library constructed by Life Technologies."

BASE COUNT 201 a 150 c 150 g 147 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 649;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 agtcacagaattatt 25
|||||
DB 517 AGTCACAGATTATT 502

RESULT 22
A0541562/c
LOCUS
DEFINITION A0541562 652 bp DNA GSS 19-MAY-1999
PC1-11-361M22.TV PC1-11 Homo sapiens genomic clone PC1-11-361M22
, DNA sequence.
ACCESSION A0541562
VERSION A0541562.1 GI:4872018
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 652)
Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter
J.C.
J.C.
Use of BAC End Sequences from Library PC1-11 for Sequence-Ready
Map Building
Unpublished (1997)
Other GSSs: PC1-11-361M22.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetlgr.org

JOURNAL
COMMENT library availability, please contact Pletier de Jong
(pleterdejong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/ordering) or from
Research Genetics (http://resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1. 652

BASE COUNT 201 a 162 c 82 g 207 t
ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 652;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaattatt 23
|||||
DB 96 CAAGTCACAGATTATT 81

RESULT 23
BF726111/c
LOCUS
DEFINITION BF726111 657 bp mRNA EST 05-JAN-2001
by01c01.y1 Human lens cDNA (un-normalized, unamplified): By Homo
sapiens cDNA clone by01c01 5', mRNA sequence.
ACCESSION BF726111
VERSION BF726111.1 GI:12042022
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 657)
Wistow, G.J., Bernstein, S., Behal, A. and Smith, D.
NEIBANK: EST analysis and bioinformatics for ocular genomics
Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@neilx.nih.gov
Plate: 01 row: c column: 01
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers
1. 657

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="by01c01"
/clone_lib="Human lens cDNA (un-normalized, unamplified):
By"
/tissue_type="lens"
/dev_stage="Adult"
/lab_host="EMD10B"
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life Technologies
, essentially following the protocols of the Superscript
Plasmid System full details of which are contained in the
manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-PGACTTATCTAGATCCGACGCGCCGCC(T)15-3']. Not I/blunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (MISC)."


```

RESULT 26
LOCUS   A0900023/c
DEFINITION
HS_3180_A2_G11_T7C CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3180 Col=22 Row=M, DNA sequence.
ACCESSION
A0900023
VERSION  A0900023.1  GI:6356213
KEYWORDS
SOURCE   GSS.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 971)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL 99380589
MEDLINE
COMMENT  Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (Info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3180 row: M column: 22
Seq primer: 17
Class: BAC ends
High quality sequence stop: 971.
FEATURES
    source
        1..971
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="Plate=3180 Col=22 Row=M"
        /clone_lib="CIT Approved Human Genomic Sperm Library D"
        /sex="male"
        /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
        E-coli DH10B"
        F-col: DH10B"
BASE COUNT      249 a      130 c      128 g      464 t
ORIGIN
Query Match      64.0%; Score 16; DB 13; Length 971;
Best local similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      10 agtcacagaataattt 25
        |||||||
Db      501 AGTCACAGAAATTATT 486

RESULT 27
LOCUS   BG871816/c
DEFINITION
602792574F1 NCI_CGAP_SG2 Mus musculus cdna clone IMAGE:4923619 5',
mRNA sequence.
ACCESSION
BG871816
VERSION  BG871816.1  GI:14222356
KEYWORDS
SOURCE   EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1318)
NIR-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

```

```

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10845 row: d column: 20
High quality sequence start: 3
High quality sequence stop: 211.
FEATURES
    source
        1..1318
        /organism="Mus musculus"
        /strain="FVB/N"
        /db_xref="taxon:10090"
        /clone="IMAGE:4923619"
        /clone_lib="NCI_CGAP_SG2"
        /lab_host="DH10B (T1 phage-resistant)"
        /note="Organ: salivary gland; Vector: pCMV-Sport6; Site:1;
        Note: Site 2: Salivary gland; Cloned unidirectionally. Primer: Oligo
        dT. Average insert size 1.3 kb. Constructed by Life
        Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      379 a      339 c      297 g      303 t
ORIGIN
Query Match      64.0%; Score 16; DB 11; Length 1318;
Best local similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 attgtcgcaatcaca 16
        |||||||
Db      1084 ATGTCGCAACTACA 1069

RESULT 28
LOCUS   BG977588
DEFINITION
PM4-C10084-040101-006-d11 C10084 Homo sapiens cdna, mRNA sequence.
ACCESSION
BG977588
VERSION  BG977588.1  GI:14380323
KEYWORDS
SOURCE   EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 172)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brenlanl,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL 20202653
MEDLINE
COMMENT  Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?pt1-PM4&t2-PM4-C10084-
040101-006-d11&t3-2001-01-04&t4-1)
Seq primer: puc 18 forward
High quality sequence stop: 165.

```

FEATURES

Source

Location/Qualifiers

1..172

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="CI0084"

/dev_stage="Adult"

/note="Organ: colon;ins: Vector: puc18; Site_1: SmaI;

Site_2: SmaI: A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT

37 a 51 c 55 g 28 t 1 others

ORIGIN

Query Match

60.0%; Score 15; DB 11; Length 172;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gtgcgaagtcacaga 18

Db 8 GTGCACAGTACACAGA 22

RESULT 29

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 30

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

within the 5' EcoRI and 3' XhoI sites. This library was

constructed by Dr. Uma Rao and Dr. David Bird at North

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

Carolina state University."

QY 8 caagtcacagaatta 22
 |||
 Db 94 CAAGTCACAGAATTA 108

RESULT 31
 BE763892 253 bp mRNA EST 19-SEP-2000
 LOCUS BE763892
 DEFINITION RC4-NT0054-120600-015-g03 NT0054 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE763892
 VERSION BE763892.1 GI:10193816
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 253)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
 JOURNAL 20202663
 MEDLINE Contact: Simpson A.J.G.
 COMMENT Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-27004922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL:
 http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC4-NT0054-120
 600-015-g03&f3=2000-06-12&tt4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 12
 Location/Qualifiers
 1..253
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0054"
 /dev_stage="Adult"
 /note="Organ: nervous_tumor; Vector: puc18; Site_1: Sma1;
 Site_2: Sma1; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 Profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 95 a 31 c 56 g 71 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 253;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaatta 22
 |||
 Db 114 CAAGTCACAGAATTA 128

RESULT 32
 AA317110 256 bp mRNA EST 19-APR-1997
 LOCUS AA317110
 DEFINITION EST19001 Lung Homo sapiens cDNA 5' end, mRNA sequence.
 ACCESSION AA317110

VERSION AA317110.1 GI:1969448
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 256)
 AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulmer,R.A., Bult
 ,C.J., Lee,N.H., Kirtress,E.F., Weinstock,K.G., Gocayne,J.D., White
 ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-wai,C., Clayton,R.A.,
 Cline,T.R., Colton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
 ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodek,A.,
 Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
 Kelley,J.C., Liu,L.-I., Marmaros,S.M., Meritck,J.M.,
 Moreno-palanges,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,L.L., Sauder,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Ulterback,T.R., Weidman,J.F., Li,X.,
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
 Dimke,D., Feng,D.-F., Ferlie,A., Fischer,C., Hastings,G.A., He,W.W.,
 Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
 Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
 Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
 ,M.R., Rosen,C.A., Haseilime,W.A., Fields,C., Fraser,C.M. and
 Venter,J.C.
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl), 3-174 (1995)

TITLE
 JOURNAL 96026280
 MEDLINE Other_ESTS: THC99418
 COMMENT Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699422
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (http://www.tigr.org/cdb/hgi/hgi.html)
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..256
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):117571"
 /db_xref="taxon:9606"
 /clone_lib="Lung"
 /dev_stage="adult"
 /note="Organ: Lung; Vector: pBluescript SK-; Site_1: EcoRI
 ; Site_2: XhoI"
 BASE COUNT 73 a 57 c 48 g 75 t 3 others
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 256;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaatta 22
 |||
 Db 189 CAAGTCACAGAATTA 203

RESULT 33
 Z45844 261 bp mRNA EST 14-NOV-1994
 LOCUS Z45844/c
 DEFINITION HSCZVE041 normalized infant brain cDNA Homo sapiens cDNA clone
 c-zve04, mRNA sequence.
 ACCESSION Z45844
 VERSION Z45844.1 GI:575078
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 261)
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 Aulfray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houigatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Picu, G., Pouliot, Y., Sebastiani-Kabakchis, C. and Tessier, A.
 IMAGE: molecular integration of the analysis of the human genome and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 JOURNAL 95277534
 MEDLINE Contact: Genethon
 COMMENT Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 Evry Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read.
 Genexpress_library_id: C; Genexpress_sequence_id: y3c-zve04
 Seq primer: (-21)M3 universal.
 FEATURES
 SOURCE Location/Qualifiers
 1..261
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="C-zve04"
 /clone_lib="normalized infant brain cDNA"
 /sex="Female"
 /tissue_type="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: lambda BA; Site_1: HindIII; Site_2: NotI; sex: Female; dev_stage=3 months old; isolate=muscular atrophy patient; tissue_type=total brain; total mRNA was oligo-(gt) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lambda BA vector. Clone library from B. Soares, Psychiatry Dept. Columbia University, USA. Normalization_method: Bento Soares, P.N.A.S. in press"
 BASE COUNT 77 a 56 c 31 g 97 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; Length 261;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 caagtcacagaatta 22
 ||||||||||||||||
 Db 138 CAAAGTCACAGATT 124
 RESULT 34
 BF745516 263 bp mRNA EST 10-JAN-2001
 LOCUS CM2-BP0826-021100-469-h05 BP0826 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF745516
 VERSION BF745516.1 GI:12072192
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 263)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Arionese, M.R., Nagai, M.A., da Silva, W. Jr., Zagro, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3451-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.

LABORATORY OF CANCER GENETICS
 LUDWIG INSTITUTE FOR CANCER RESEARCH
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM2&t2=CM2-BP0826-021100-469-h05&t3=2000-11-02&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 129
 High quality sequence stop: 263.
 FEATURES
 SOURCE Location/Qualifiers
 1..263
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BP0826"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 77 a 55 c 60 g 71 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; Length 263;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 gcaagtcacagaatt 21
 ||||||||||||||||
 Db 217 GCAAGTCACAGATT 231
 RESULT 35
 AA115083 266 bp mRNA EST 14-MAY-1997
 LOCUS 2106C10.T1 Soares_pregnant_uterus_NbHPV Homo sapiens cDNA clone
 DEFINITION IMAGE:491538 5', mRNA sequence.
 ACCESSION AA115083
 VERSION AA115083.1 GI:1670318
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 265)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chippelli, B., Chissole, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Martis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellendberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevisakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 Contact: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNC; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 555 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amerasham.
 Location/Qualifiers

source

```

1. 266
/organism="Homo sapiens"
/db_xref="GDB:3806125"
/db_xref="taxon:9606"
/clone="IMAGE:491538"
/clone_lib="Soares_pregnant_uterus_NbHpu"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(40) primer [5', AACTGGAAGAATTCGCCGCCCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bernaldo."
BASE COUNT      61 a      67 c      64 g      69 t
ORIGIN

```

```

Query Match      60.0%; Score 15; DB 10; Length 266;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 gtcacagaattatt 25
|||||
Db 91 gtcacagaattatt 77

```

```

RESULT 36
BF800965      282 bp      mRNA      EST      12-JAN-2001
LOCUS      QV4-C10100-231000-457-d12 C10100 Homo sapiens cDNA, mRNA sequence.
DEFINITION      BF800965
ACCESSION      BF800965
VERSION      BF800965.1 GI:12129954
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 282)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1-QV4&t2-QV4-C10100-
231000-457-d12&t3=2000-10-23&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 59
High quality sequence stop: 282.

```

```

FEATURES
Source
1. 282
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="C10100"

```

```

/dev_stage="Adult"
/notes="Organ: colon lns; Vector: puc18; Site_1: SmaI; Site_2: SmaI. A mini-library was made by cloning products derived from ORESTES PCR (0.5. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      68 a      75 c      67 g      72 t
ORIGIN

```

```

Query Match      60.0%; Score 15; DB 11; Length 282;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtgcgaatcacc 15
|||||
Db 51 atgtgcgaatcacc 37

```

```

RESULT 37
AA344205      286 bp      mRNA      EST      21-APR-1997
LOCUS      AA344205
DEFINITION      EST50086 Gall bladder I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION      AA344205
VERSION      AA344205.1 GI:1996443
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 286)

```

```

REFERENCE
AUTHORS      Adams,M.D., Kerlavage,A.R., Fleischman,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Well,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Guelm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Peng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.M.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G., L., Rubin,S.M., Dillion,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Other-ESTs: THC99418
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavage@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.cigr.org/ldb/hgi/hgi.html)
Seq primer: M13 Reverse.

```

```

FEATURES
Source
1. 286
/organism="Homo sapiens"
/db_xref="ATCC (lnost):145738"
/db_xref="taxon:9606"
/clone_lib="Gall bladder I"
/sex="female"

```

Query Match 60.0%; Score 15; DB 10; Length 286;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 8 caagtcacagaatta 22
|||||

Db 197 CAAGTCACAGAATTA 211

RESULT 38
B53227/c 292 bp DNA GSS 20-JUN-1998
LOCUS C1T-HSP-2009H2.TK C1T-HSP Homo sapiens genomic clone 2009H2, DNA
DEFINITION
SEQUENCE.
ACCESSION B53227
VERSION B53227.1 GI:2607561
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 292)
Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
and Verter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
Unpublished (1997)
JOURNAL Contact: Mark Adams
COMMENT Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@ligr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.ligr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..292
/organism="Homo sapiens"
/db_xref="GDB:704081.2"
/db_xref="taxon:9606"
/clone_id="2009H2"
/clone_lib="C1T-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC1; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 58 a 64 c 90 g 80 t

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 8 caagtcacagaatta 22
|||||

Db 248 CAAGTCACAGAATTA 214

RESULT 39
BF743951/c

LOCUS BF743951 317 bp mRNA EST 10-JAN-2001
DEFINITION CM2-BF0826-181000-428-BF0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF743951
VERSION BF743951.1 GI:12070627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 317)
Dias,Neto,E., Garcia Correia,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the RAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?PL=CM2&t2=CM2-BF0826-181000-428-c02&t3=2000-10-18&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 317.

FEATURES
source Location/Qualifiers
1..317
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="BF0826"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (O.S. Letters Patent Application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 96 a 71 c 49 g 101 t

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 7 gcaagtcacagaatt 21
|||||

Db 45 GCAAGTCACAGAATP 31

RESULT 40
BH121038/c 324 bp DNA GSS 19-JUL-2001
LOCUS BH121038
DEFINITION RPCT-24-230E13.TV RPCT-24 Mus musculus genomic clone RPCT-24-230E13
DNA sequence.
ACCESSION BH121038
VERSION BH121038.1 GI:14964550
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 324)
 AUTHORS Zhao,S., Niemman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
 Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregziorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-24
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: RPCI-24-230E13.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 230 row: E column: 13
 Seq primer: T7
 Class: BAC ends.

FEATURES
 Location/Qualifiers
 1..324
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-230E13"
 /clone_1lb="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: PTARBAC1. Site 1: BamHI. Site 2: BamHI;
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the PTARBAC1 cloning vector at the
 BamHI sites using Mbol partially digested male C57BL/6J
 DNA."

BASE COUNT 93 a 57 c 88 g 86 t

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 324;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaatt 21
 |||||

Db 26 GCAAGTCACAGAATT 12

RESULT 41
 T30901/c 327 bp mRNA EST 06-SEP-1995
 LOCUS EST24536 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA
 DEFINITION
 sequence.
 T30901
 ACCESSION T30901 GI:612999
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 327)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulton,R.A., Bult
 ,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
 ,O., Sutton,G., Blake,J.A., Brandon,R.C., Chin,M.-W., Clayton,R.A.,
 Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
 ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A.,
 Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S., Kelley,J.M.,
 Klinek,K.M., Kelley,J.C., Liu,L.-T., Marmaros,S.M., Merrick,J.M.,
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
 Bedharik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,

Dimke,D., Feng,P., Ferlie,A., Fischer,C., Hastings,G.A., He,W.-W.,
 Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L.,
 Kunsch,C., Ji,H., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei
 ,Y.-F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
 ,M.R., Kosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
 Venter,J.C.
 TITLE Initial Assessment of Human Gene Diversity and Expression Patterns
 JOURNAL Nature 377, 3-174 (1995)
 MEDLINE 96026280
 COMMENT Other_ESTs: THC10632
 Contact: Venter, JC
 The Institute for Genomic Research
 932 Clopper Rd, Gaithersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: tdbinfo@tdb.tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please contact the TIGR Database
 (tdbinfo@tdb.tigr.org)
 Seq primer: MJ3 Reverse.

FEATURES
 Location/Qualifiers
 1..327
 /organism="Homo sapiens"
 /db_xref="ATCC (Inhost):100882"
 /db_xref="taxon:9606"
 /clone_1lb="Human Brain"
 /note="Organ: brain"

BASE COUNT 89 a 75 c 47 g 115 t 1 others

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaatta 22
 |||||

Db 295 CAAGTCACAGAATTA 281

RESULT 42
 BF743964/c 335 bp mRNA EST 10-JAN-2001
 LOCUS BF743964
 DEFINITION CM2-BF0826-181000-428-h04 BF0826 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF743964
 VERSION BF743964.1 GI:12070640
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 335)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zagro,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunsstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-27049922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FARESP/TIGR Human Cancer Genome
 Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=CM2&l2=CM2-BT0826-181000-428-h04&l3=2000-10-18&l4=1)

Seq primer: puc 18 forward
High quality sequence start: 40
High quality sequence stop: 335.

FEATURES

source

```

1..335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0826"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      101 a      76 c      55 g      103 t
ORIGIN

```

Query Match

Best local Similarity 100.0%; Score 15; DB 11; Length 335;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaatt 21
|||||

Db 61 GCAAGTCACAGAATT 47

RESULT 43

```

BF743954/c 346 bp mRNA EST 10-JAN-2001
LOCUS      CM2-BT0826-181000-428-d07 BT0826 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF743954
ACCESSION  BF743954
VERSION    BF743954.1 GI:12070630
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
            Brunstein,A., deoliveira,P.S., bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

```

REFERENCE

AUTHORS

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=CM2&l2=CM2-BT0826-181000-428-d07&l3=2000-10-18&l4=1)

Seq primer: puc 18 forward
High quality sequence start: 43
High quality sequence stop: 346.

FEATURES

source

```

1..346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0826"

```

```

/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      103 a      78 c      54 g      111 t
ORIGIN

```

Query Match

Best local Similarity 100.0%; Score 15; DB 11; Length 346;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaatt 21
|||||

Db 73 GCAAGTCACAGAATT 59

RESULT 44

```

T07051      370 bp mRNA EST 30-JUN-1993
LOCUS      EST04940 Fetal brain, Striatogene (cat#936206) Homo sapiens cDNA
DEFINITION clone HFBEC79, mRNA sequence.
ACCESSION  T07051
VERSION    T07051.1 GI:318200
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
            3,400 expressed sequence tags identify diversity of transcripts
            from human brain
            Nature Genet. 4, 256-267 (1993)

```

REFERENCE

AUTHORS

TITLE

3,400 expressed sequence tags identify diversity of transcripts from human brain

JOURNAL

Nature Genet. 4, 256-267 (1993)

COMMENT

Contact: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13-21.

FEATURES

Location/Qualifiers

```

1..370
/organism="Homo sapiens"
/db_xref="ATCC (inhost):83698"
/db_xref="taxon:9606"
/clone_lib="HFBEC79"
/clone_lib="Fetal brain, Striatogene (cat#936206)"
/note="Vector: LambdaZAP-II; 17-18 wk gestation, female; oligo-dr + random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb average insert size."
BASE COUNT      137 a      58 c      68 g      106 t
ORIGIN

```

Query Match

Best local Similarity 100.0%; Score 15; DB 11; Length 370;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaatta 22
|||||

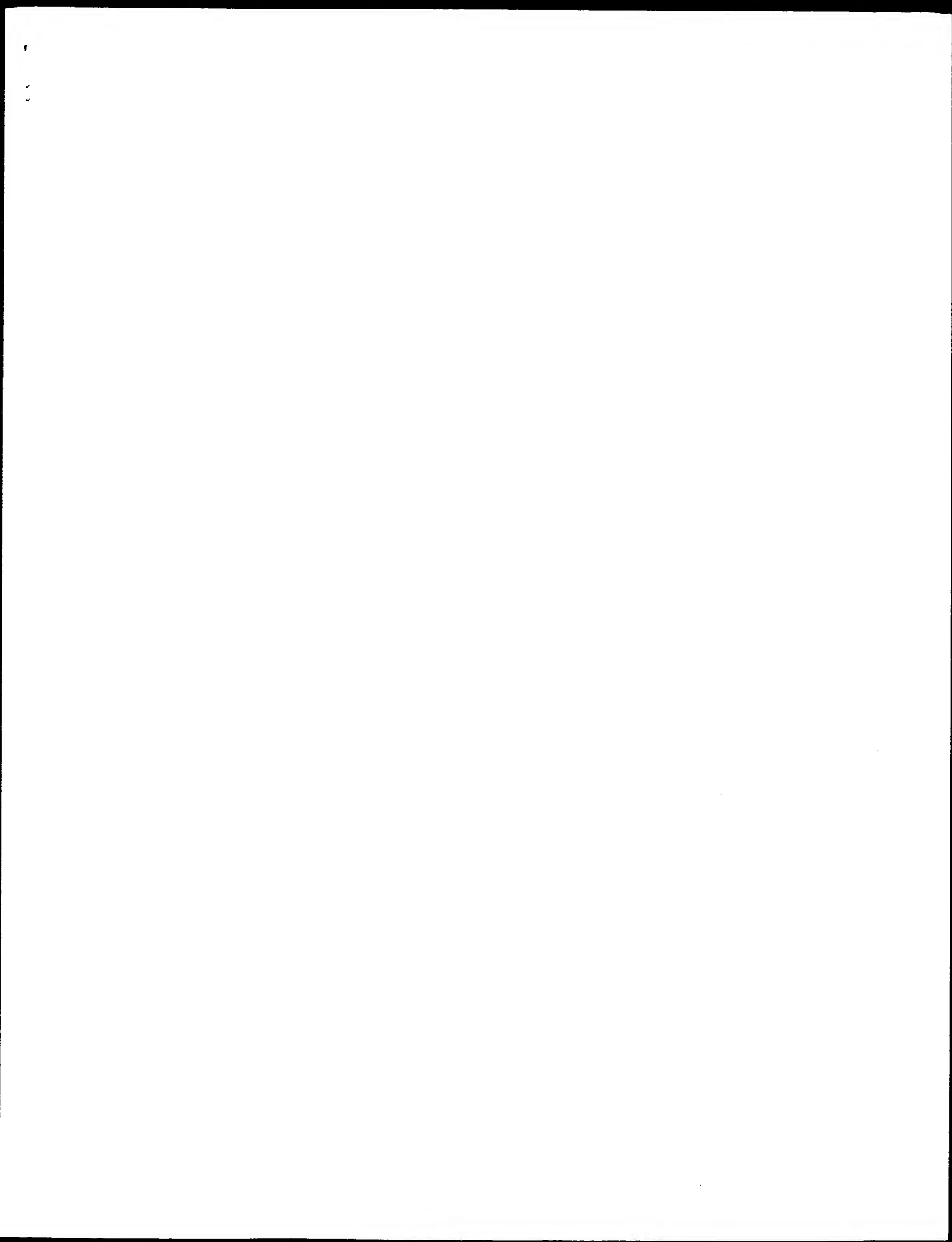
Db 113 CAAGTCACAGAATTA 127

RESULT 45

```

H53286/c 381 bp mRNA EST 20-SEP-1995
LOCUS      YG85f09.r1 Soares fetal liver spleen INF1S Homo sapiens cDNA clone
DEFINITION

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:50:42 : Search time 1757.07 Seconds
(Without alignments) 234.726 Million cell updates/sec

Title: US-09-396-196f-3

Perfect score: 1 gttgtgacatcgtcgttgaagcg 25

Sequence: 1 gttgtgacatcgtcgttgaagcg 25

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 270703

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_btg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pal:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_btgo_hum:*
31: em_btgo_inv:*
32: em_btgo_rod:*
33: em_btgo_hum:*
34: em_btgo_inv:*
35: em_btgo_rod:*
36: em_btgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	128	1	ECOB10B	M27731 Escherichia
2	25	100.0	1041	6	AR029499	AR029499 Sequence
3	25	100.0	1041	6	AR034916	AR034916 Sequence
4	25	100.0	1084	6	AL11530	AL11530 B108 gene o
5	25	100.0	5526	1	AF250776	AF250776 Unculture
6	25	100.0	5793	1	ECOB10	J04423 E.coli 7.8-
7	25	100.0	5872	6	A38246	A38246 Sequence 1
8	25	100.0	5872	6	A38251	A38251 Sequence 6
9	25	100.0	5872	6	A93674	A93674 Sequence 1
10	25	100.0	5872	6	A93679	A93679 Sequence 6
11	25	100.0	5872	6	AR101809	AR101809 Sequence
12	25	100.0	5872	6	AR101810	AR101810 Sequence
13	25	100.0	11022	1	AE000180	AE000180 Escherich
14	25	100.0	13501	1	AE005258	AE005258 Escherich
15	25	100.0	13501	1	AE002553	AE002553 Escherich
16	24	96.0	1121	6	AE00893	AE00893 Genomic DNA
17	20	80.0	297816	1	AE004192	AE004192 Vibrio ch
18	17	68.0	383	3	S64534	S64534 Drosophila
19	17	68.0	461	3	S64532	S64532 Drosophila
20	17	68.0	516	3	DMDECG1B	X76231 D.melanog
21	17	68.0	539	3	S64530	S64530 Drosophila
22	17	68.0	539	3	S64531	S64531 Drosophila
23	17	68.0	617	3	S64529	S64529 Drosophila
24	17	68.0	774	3	DSDECI1R6	X96931 D.secellia
25	17	68.0	852	3	DSDECI1R2	X96927 D.simulans
26	17	68.0	852	3	DSDECI1R3	X96928 D.simulans
27	17	68.0	930	3	DMDECI1R1	X96926 D.melanog
28	17	68.0	3741	3	DR0DEC1A	M35887 Drosophila
29	17	68.0	3755	3	DR0DEC1B	M35888 Drosophila
30	17	68.0	5697	3	DR0DEC1C	M35889 Drosophila
31	17	68.0	33204	2	AC017995	AC017995 Drosophila
32	17	68.0	161340	2	AC016952	AC016952 Homo sapi
33	17	68.0	306267	3	AE003442	AE003442 Drosophila
34	16	64.0	774	3	DMDECI1R5	X96930 D.mauritian
35	16	64.0	1044	5	AF250347	AF250347 Xenopus 1
36	16	64.0	1614	5	AF281080	AF281080 Xenopus 1
37	16	64.0	5000	8	AF120145	AF120145 Sequence
38	16	64.0	5000	8	AF13940	AF13940 Arabidops
39	16	64.0	6887	8	YL1012084	AJ012084 Yarrowia
40	16	64.0	11328	1	AE004322	AE004322 Vibrio ch
41	16	64.0	13856	1	AE004026	AE004026 Xylella f
42	16	64.0	15247	1	AE004580	AE004580 Pseudomon
43	16	64.0	31557	9	HS31464	Z69667 Human DNA s
44	16	64.0	41200	8	SC8520X	Z49705 S.cerevisia
45	16	64.0	46340	6	AX008043	AX008043 Sequence
46	16	64.0	62927	8	AB018119	AB018119 Arabidops
47	16	64.0	65679	2	AC013956	AC013956 Drosophila
48	16	64.0	74127	2	AC020125	AC020125 Drosophila
49	16	64.0	78529	2	AC012807	AC012807 Drosophila
50	16	64.0	79048	2	AC008298	AC008298 Arabidops
51	16	64.0	102540	8	AC079281	AC079281 Arabidops
52	16	64.0	109683	2	AC073020	AC073020 Homo sapi
53	16	64.0	110000	2	LMFLCHR31_15	Continuation (16 o
54	16	64.0	150594	8	AP001859	AP001859 Oryza sat
55	16	64.0	156425	8	AP002897	AP002897 Oryza sat
56	16	64.0	158148	2	AC015463	AC015463 Homo sapi
57	16	64.0	158560	2	AC013357	AC013357 Homo sapi
58	16	64.0	166919	2	AL591675	AL591675 Mus muscu
59	16	64.0	167862	9	AC011966	AC011966 Homo sapi
60	16	64.0	168177	2	AC023914	AC023914 Homo sapi
61	16	64.0	168893	2	AC007532	AC007532 Drosophila
62	16	64.0	171069	2	AP003845	AP003845 Oryza sat
63	16	64.0	177163	9	HS134E15	AL022067 Human DNA
64	16	64.0	177857	3	AC010211	AC010211 Drosophila
65	16	64.0	177989	2	AL591430	AL591430 Mus muscu
66	16	64.0	187272	2	AC069076	AC069076 Homo sapi
67	16	64.0	188055	2	AC023593	AC023593 Homo sapi
68	16	64.0	188156	2	AC079277	AC079277 Mus muscu
69	16	64.0	189252	2	AC018930	AC018930 Homo sapi
70	16	64.0	190029	2	AC068664	AC068664 Mus muscu

SUMMARIES

```

71 16 64.0 191677 2 AC090534 Mus muscu
72 16 64.0 200000 2 AC007104 Homo sapi
73 16 64.0 200900 2 AC013410 Homo sapi
74 16 64.0 214241 2 AC084163 Mus muscu
75 16 64.0 221420 2 AC011822 Homo sapi
76 16 64.0 228155 2 AC018755 Homo sapi
77 16 64.0 260418 2 AC006463 Homo sapi
78 16 64.0 266344 2 AC005158 Homo sapi
79 16 64.0 277023 3 AC003519 Drosophi
80 16 64.0 291545 3 AC003600 Drosophi
81 15 60.0 306 14 PHE7D D50348 Human papil
82 15 60.0 358 11 G00476 fruit fly S
83 15 60.0 503 3 DEDGCG1B X76233 D.erecia de
84 15 60.0 708 6 A86358 A86338 Sequence 99
85 15 60.0 708 6 AK155831 AR155831 Sequence
86 15 60.0 708 6 E66356 E66356 Genome DNA
87 15 60.0 1041 1 EHHIOTOP2 U38648 Erwinia her
88 15 60.0 1172 8 ZMSNORNA2 V11049 2. mays smal
89 15 60.0 1616 10 AF239157 AF239157 Rattus no
90 15 60.0 1623 10 AF009246 AF009246 Mus muscu
91 15 60.0 1683 6 A72446 A72446 Sequence 33
92 15 60.0 1683 6 A80156 A80156 Sequence 33
93 15 60.0 2018 3 AE002676 AE002676 Drosophi
94 15 60.0 2021 2 AC019735 AC019735 Drosophi
95 15 60.0 2546 9 AK026448 AK026448 Homo sapi
96 15 60.0 3569 3 CEU20168 U20168 Caenorhabdi
97 15 60.0 4116 1 PSEKDG0H M69160 Pseudomonas
98 15 60.0 5607 3 DROPROSA M69160 Drosophila
99 15 60.0 5869 9 AF181672 AF181672 Homo sapi
100 15 60.0 6017 3 DMPROSPER Z11743 D. melanoga

```

ALIGNMENTS

```

RESULT 1
ECOBIOB 128 bp DNA BCT 03-JAN-1995
LOCUS Escherichia coli biotin (biob) gene, early terminator region.
DEFINITION M27731.1 GI:341755
ACCESSION
VERSION
KEYWORDS biob gene; biotin.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 128)
AUTHORS Nath.S.K.
TITLE Attenuation of transcription of biotin genes in Escherichia coli
JOURNAL Can. J. Microbiol. 34 (12), 1288-1296 (1988)
MEDLINE 89167942
FEATURES
source location/Qualifiers
1..128
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
gene 4..128
/organism="biob"
-10_signal 4..9
/organism="biob"
mRNA 20..>128
/organism="biob"
terminator 103..121
/organism="biob"
/organism="early terminator"
BASE COUNT 31 a 28 c 34 g 35 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gtgcgtgacatcgtcttgaagc 25
|||||
DB 18 GTTGCTGATCTGCTGTTGAAGC 42

```

```

RESULT 2
LOCUS AR029499 1041 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES Location/Qualifiers
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gtgcgtgacatcgtcttgaagc 25
|||||
DB 57 GTTGCTGATCTGCTGTTGAAGC 81

```

```

RESULT 3
LOCUS AR034916 1041 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES Location/Qualifiers
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gtgcgtgacatcgtcttgaagc 25
|||||
DB 57 GTTGCTGATCTGCTGTTGAAGC 81

```

```

RESULT 4
LOCUS A11530 1084 bp DNA PAT 09-FEB-1994
DEFINITION Biob gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.

```



```

ORGANISM      Escherichia coli
REFERENCE      Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
AUTHORS        Escherichia
JOURNAL         1 (bases 1 to 1084)
FEATURES       Patent: GB 2216530-A 16 11-OCT-1989;
source         Location/Qualifiers
               1..1084
               /organism="Escherichia coli"
               /db_xref="taxon:562"
gene           24..1064
               /gene="BIOB"
               /gene="BIOB"
               /gene="BIOB"
CDS            24..1064
               /gene="BIOB"
               /codon_start=1
               /transl_table=11
               /protein_id="CA00965.1"
               /db_xref="GI:490219"
               /db_xref="SWISS-PROT:P12996"
               /translation="MAHPRMRLTQVTELFKPLDLLFEAOQVHROHDPPOVYST
               LLSIKTGACPECKCYCROSSRYKTGLEALMEVROVLESARAKAAGSTRCMGAAM
               KNHERDMPTLEBOWGVKAMGLEACHTLGLTSESQORLANGLDYNNHLDTSPEP
               VGNITTRTYOERLDLTKERVDAGIKVCGSGIVGLGETVKDRAGLLQLANLTPES
               VPINMLVKVKGTPLANDVDVADFTRTAVARIMPTSYVRLSAGREOMECOTQAMC
               FMAGNSIFYGCKLTTTPNEPBEKDLQFRKRLGLNPQTAVLAGDNEQOORLEQALMT
               PDIDEVYNAAL"
BASE COUNT    271 a      286 c      318 g      209 t
ORIGIN
Query Match   100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY            1 gttgctgacatcgtctgttaagca 25
               |||||||
Db            80 GTTGCTGATCTGCTGTTGAAGCG 104
RESULT        5
LOCUS         AF250776      5526 bp      DNA      BCT      31-JAN-2001
DEFINITION    Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
               modc-bioA intergenic region, DAPA-aminotransferase bioA (bioA),
               biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
               biotin biosynthesis protein BioC (bioC) genes, complete cds; and
               dehydrobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION     AF250776
VERSION       AF250776.1 GI:12620124
KEYWORDS
SOURCE        uncultured bacterium pCosHE2.
               Bacteria: environmental samples.
REFERENCE     1 (bases 1 to 5526)
AUTHORS       Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Strelt, W.R.
TITLE         Direct cloning from enrichment cultures, a reliable strategy for
               isolation of complete operons and genes from microbial consortia
               Appl. Environ. Microbiol. 67 (1), 85-99 (2001)
JOURNAL       20575196
MEDLINE       1113432
PUBMED        2 (bases 1 to 5526)
REFERENCE     Entcheva, P., Liebl, W. and Strelt, W.R.
AUTHORS       Direct Submission
TITLE         Submitted (31-MAR-2000) Mikrobiologie und Genetik. Universitaet
               Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
JOURNAL       Goettingen
FEATURES       Location/Qualifiers
               1..5526
               /organism="uncultured bacterium pCosHE2"
               /db_xref="taxon:143797"
               /clone="pCosHE2"
               /note="unknown organism, cosmid clone derived from
               environmental consortium"
               complement(52..528)
CDS

```

```

/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kDa protein in modc-bioA
intergenic region
protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDGDLPKRRHVENGCYGDGNISPLHADDVPAGTK
SEVVTGYPDAPGCGSSGMMWVYVNLADPTRYLPDGGSGVYAMPDGYLQTRTPGKAG
YDGAAPRGKTHRYIFPYHALDIERIDVDEACAMGPNVHFHSLASATIMFS"
complement(587..1876)
/gene="BioA"
complement(587..1876)
/note="7,8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase BioA"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDDLAFPDORHIMHPTSMTPLPVYPVVSARGCCELLISDGR
LVDCSSWMAIHCNTNHPOLAAKMSQIDAMSHVFCGITHAPALIELCKLIVAMPFP
LKCVFLADSGSVAVEAVAKMALQYWAQGEARQRFLLFRNGYHDDTEGAMSVDPPDS
MHSLLKGYLPENLFAPOSERMDGEMDERDVGARLMAARHBIAAVILPEIVQAG
GMRMYHPMLKRIKICDREGILLIADIATGFTGKLFACBAEATAPILCLGKAL
TGCMTLSATLITTRVAVETISNGEAGCFMGHPTEGPNFLACANAASALIESGDMQO
OYADIEVOLREGIAPARDAENYADYRVIGALGVETTHPVMAALOKFVDSQVHIR
FKLIVLMPPTYLTPQOLRLTAANVRVQDETFQO"
1963..3003
/gene="BioB"
1963..3003
/gene="BioB"
/gene="BioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHPRMRLTQVTELFKPLDLLFEAOQVHROHDPPOVYST
LLSIKTGACPECKCYCROSSRYKTGLEALMEVROVLESARAKAAGSTRCMGAAM
KNHERDMPTLEBOWGVKAMGLEACHTLGLTSESQORLANGLDYNNHLDTSPEP
VGNITTRTYOERLDLTKERVDAGIKVCGSGIVGLGETVKDRAGLLQLANLTPES
VPINMLVKVKGTPLANDVDVADFTRTAVARIMPTSYVRLSAGREOMECOTQAMC
FMAGNSIFYGCKLTTTPNEPBEKDLQFRKRLGLNPQTAVLAGDNEQOORLEQALMT
PDIDEVYNAAL"
3000..4154
/gene="BioF"
3000..4154
/gene="BioF"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase BioF"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSWQEKINAALDARRADALRRRYPVAGAGRWLVADROYLNF
SENDYIGLSHHPQIIRAMQOGEQFSGISGSGSVYHQALESALSPORRAHNDVHSLA
ILFIFGEANQVIAAMAKKEDRIADRLSHSILSALSPORRAHNDVHSLA
LLASPCGQQLVYTEGVSFMSDDSNAPLAEIQOYVQNNHMLKLVDAHATGYIGERG
SCWLOKVRPELLVTFEGFVSAAVLCSTVADYLIQFAHLLYTSMPKPAQAL
RASLAVIRNDEGDARERKLVSLIAHFRAGVODLPPTLADSCSAIOLPLIVGNRQAL
AEKLRQGCWYAIRPPVPAGTARLRULTFAHMODIDRLLEVLHGNG"
4141..4896
/gene="BioC"
4141..4896
/gene="BioC"
/note="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein BioC"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVNRKQAIAAFGRAALYEQHADLQROSADVILLAMLPKRYT

```

[illegible]

BASE COUNT 1363 a 1554 c 1631 g 1245 t
 ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 1; Length 5793;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtcgtgacatcgtctgttgaagcg 25
 Db 2068 GTTCTGATCTGCTGTTGAAGCG 2092

RESULT 7
 A38246 5872 bp DNA PAT 05-MAR-1997
 LOCUS
 DEFINITION Sequence 1 from Patent WO9408023.
 A38246
 VERSION A38246.1 GI:2294844
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 JOURNAL Patent: WO 9408023-A 1 14-APR-1994;
 LONZA AG (CH)

COMMENT
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331
 Other publication JP 85016947 960227.
 location/Qualifiers

FEATURES
 source
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A-15/9"
 1..96
 /function="PROMOTER PTAC"
 /evidence=experimental
 23..28
 45..49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental
 105..119
 /standard_name="BIOR RBS NO.9"
 /evidence=experimental
 117..1157
 /gene="BIOB"
 117..1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /number=1
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CAA02325.1"
 /db_xref="GI:2294845"
 /translation="MAHRPRWTLISOVTELFEPKLLDLEFAOVVHROHPROVYST
 LLSKTGACPEDEKTCPOSSRYKGLFAERLMEVDVLESARAKAGSTRPGGAAM
 KNPHRMPYLEOVGVKAMGLEACMTIGTSLSSQAKLANGLDYNNINLPTSPF
 YGNLITTPQERLDLTKVRDAGIKVCSGIVGLGTVKDRAGLLQLANLPTSPF
 VPIMLVKVGKTPLEADNDVDADFIRTIIVARIIMPSTVYRLSAGREONEDTOAMC
 FMAGANSIFYGCKLITPPNEEDKDLDFPKLGINPQFVAVLAGDNDQQRRLDQALMT
 PDTPDYNNAAAL"
 2284..2297
 /standard_name="BIOC RBS"

gene
 2295..3050
 /gene="BIOC"
 2295..3050
 /gene="BIOC"
 /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=3
 /product="PROTEIN"
 /protein_id="CAA02326.1"
 /db_xref="GI:2294846"
 /translation="MATVKNQALAAAFGRAAHYEONHADLQROSDALLAMLPOKRYT
 HYLDAGCGPGGMSRHREHNOVATLDSPEMLVOAFGDIAGDIESTPLATA
 TFDLAMSNLAVOMCGNLSTALRELYRVVRPGVAVFTTLVCGSLPELHQAOVDERP
 HANRLPPDEIEOSIGNVYOHIOPLITLMPEDDALSAMRSLKIGATHLHEGRDPRIL
 TRSQRLQALAMPQOGKRYPLTHLFLGLVARE"
 3742..3752
 /standard_name="BIOA RBS"
 3750..5039
 /gene="BIOA"
 3750..5039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE:
 8-AMINO-7-OXONONANOATE AMINOTRANSF."
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /number=5
 /evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAA02327.1"
 /db_xref="GI:2294847"
 /translation="MTTDLAFPDORHHPYTSMTSPVYPVVSAEGCELISDGR
 LVDGSSMAALHGYNHNOLNANMSQIDASHVMFGITPAALIELCRILVAMP
 LECVPLADSGSYAVVPAKRALQYQAKGEARORLTPRNGYHDTGCMSCPDPS
 MSLMKGYLPENLFAAPQSRMDGEMDERDWGEARLMAHREHAEIATIEPTVOGAG
 GMRWHHPMLKRIKICDEGLILADEIAGFGTKLFACEHAEIAPDILCYGKAL
 TGGTITLSATLTTRVAETIISNGEGCFMHPGTFGNPLACAAANASLALIESGDWO
 OVADIEVOLREOLAPARDAEMVADRVGAGISGVETTPVMAALOKFFVBOGVIRP
 FGLIYIMPYITLPQQLRLTAANVRAVDETFECQ"
 5088..5100
 /standard_name="ORFI RBS"
 5098..5574
 /gene="ORFI"
 5098..5574
 /gene="ORFI"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=6
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAA02328.1"
 /db_xref="GI:2294848"
 /translation="MKLISNDLRDGDKLPHRHVENGMGYDGDNISPHLADVDYAGTK
 SEVVTCPDADAPGSGMWHVYNLPADTRVLPQFSGGLVAMPDGVLOTPTDYGKTS
 YDGAAPKGETHRYITFVHALDIERIDVDEGASGAMGFNVHFSIASASITAMS"
 5583..5605
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtcgtgacatcgtctgttgaagcg 25
 Db 173 GTTCTGATCTGCTGTTGAAGCG 197

```

RESULT      8
LOCUS       A38251      5872 bp      DNA
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION      A38251.1  GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE       BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     other publication PL 308301 950724
            other publication CA 2145400 940414
            other publication AU 4820293 940426
            other publication HU 71781 960228
            other publication SK 42095 951108
            other publication CZ 9500809 950913
            other publication FI 951547 950331
            other publication JP 85016947 960227.
FEATURES
    source
        1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A15-9"
            1141..1156
            /standard_name="B10F RBS"
            1154..22308
            /gene="B10F"
            1154..2308
            /gene="B10F"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /number=2
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAA02329.1"
            /db_xref="GI:2294850"
            /translation="MSWDERINAAALDARRARADALRRRYVAGAGRWLVADDDROYLNF
            SSNDYLGLSHHHPQIIRAMQCAEQFGISGSGSHVGSVHQALEEELAEWLGYSRA
            LFIISGFANQAVITAAAMAKEDRIADRLSHASLSLEASLSPQLRFAHNDYPLAK
            LLASPCGQOMVTEGVFSMDGDSAPLAEIQOYTOOHGMLWVDAHGTVGBOGRG
            SCWLKVRPELLVVTGKGFGVGAVALVCSSTVADYLLQFAHLLIYSTSMPPAQOAL
            RASLAVIRSDGGRAREKLAALITFRACVQDPLTLADSCSAIOPLVIGDSRALDL
            AEKLRQGGCVTAIRPPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNG"
            3030..3045
            /standard_name="B10D RBS"
            3043..3753
            /gene="B10D"
            3043..3753
            /gene="B10D"
            /standard_name="DETHIOBIOTIN SYNTHASE"
            /EC_number="6.3.3.3"
            /codon_start=1
            /transl_table=11
            /number=4
            /evidence=experimental
            /product="DTB SYNTHASE"
            /protein_id="CAA02330.1"
            /db_xref="GI:2294851"
            /translation="MSKRYFVGTDTVEGKTIVASCALLQAAKAGRTAGYKPVYAGS
            EKTPEIRNSDALAIORNSQLDVAITNPYFAEPTSPHIIISAGEGRPTISIVMSAS
            LBALEQADWMLVEGAGGFWPLSDTPTFAPMVYQEOPLVILVGVKIGCINHAMLTA
            OVIOHAGLITLACWANDVTPPKRRAAEVMTITLTKMIPAPLLGETFWLAENPENATCK
            YINLAFVDPASTLGTPTSL"

```

```

BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 gttcctgatactcgtgttgaacg 25
        |||
Db      173 GTTGCTGATCTGCTGTTGAGCG 197

RESULT      9
LOCUS       A93674      5872 bp      DNA
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION      A93674.1  GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O. and Brass, J.
TITLE       Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES
    source
        1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..49
            /standard_name="PROMOTER PTAC"
            /evidence=experimental
            105..119
            /standard_name="B10B RBS NO. 9"
            /evidence=experimental
            117..1157
            /gene="B10B"
            117..1157
            /gene="B10B"
            /codon_start=1
            /transl_table=11
            /evidence=experimental
            /product="B10F IN SYNTHASE"
            /protein_id="CAB69590.1"
            /db_xref="GI:6741863"
            /translation="MAHRRPWITTSQVTELFERKPLDLLEFAQVYHROHEDPROVOYST
            LLSKTGACPEDCKYCPQSSRYRTGLFAERLMEVOYLESARRAKAGSTRFCMGAW
            KNPERMPYLEQWVGVKAMGLEACWTTLTLESQORLANAGLDYVYNNHNDTSPDF
            YGNLIITRTYOEERLDITLEKVDAGIKVSGSGLVGETVDRAGLLLOLANLPTPES
            VPINMLKVGKGTPLADNDVDVADFIFTIIVARIMPTSVYRLSAGREDMBOOTAMC
            EMAGANSIFYGCKKILTPNPEEDKDLQLEFRKGLNLNQAVIAGNDQQRILEQALMT
            PDIDEEYNAAL"
            2284..2297
            /standard_name="B10C RBS"
            2295..3050
            /gene="B10C"
            2295..3050
            /gene="B10C"
            /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
            /codon_start=1
            /transl_table=11
            /product="PROTEIN"
            /protein_id="CAB69591.1"

```

```

/bb_xref="GI:6741864"
/translation="MATVKNQAI AAFGRAAHVEQHADLQKOSADALAMLPQRKYT
HTVADSGGMSRRHRRHQAQVATLDSPLVVOARQKADADHLAGIESTLPATA
TFDLASNLAVOMCGNLSTALRELYVVPKGVAFPTLVQSLPELHOAMQAVDERP
HANRLPPEIOLQSLNGVHYQHIIOPITTFPDDALSAMSLGICATHTLHEGRDRIL
TTSQLOLQWAMPQDGRPLTYHLEGLVARE"
RBS
3742..3752
/standard_name="BIOA RBS"
gene
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANONATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/translation=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDORHIMHPYTSMTSPLPVVPVSAEGCELLSDGR
LVDSMSMAAIIHGVNHPOLNMAKSQIDAMSHVMFGITTHAPALIECRKIVAMPPOP
LRCVPLADSGVAVEYAKMALQYQAKKEARQRELTTRNGTHGDTFGMSVCDPDS
MHSVMKGYLPENLFAPQSRMDGEWDERDWGEPALMAAHHEHTAAVITEPIVQAG
GMRMHPEMLKRLRKTQREGLILADELATQFRTGKLFEACAEIADILCLKAL
TGGTMTLSATLTREVAETISNGEAGCFMHPGTFMGKPLACAAASIALILSCMOQ
QVADIEVOLREOLAPARDAENVADVGLAIGVETTHPVNMAALQKFEVEGVWIRP
FGKILMPPIYILLPOQLRLTAANRAVQDETFPCQ"
RBS
5088..5100
/standard_name="ORFI RBS"
gene
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/translation=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDCKLPFRHVENGMGYDGNISPHILAMDDVPAGTK
STVVTCTPDADPFGSGMHWVYVNLPAIDRVLPQSGSLVAMPDGVLOTRTDFKGT
YDGAAPKGETHRYITVHALDIERIDVEGASGAMVGFNVHFSIASASLTAMS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gttcgtgacatcgtctgttgaagcg 25
|||||
Db 173 gttcgtgacatcgtctgttgaagcg 197

```

```

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
gene
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/translation=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAALDARRADALRRRYPVAGAGRWLVADDBOYINF
SSNDYLGSHHPOLIRAMQOAGEQEGISGSGSVGSYVHOALEBELAEMIGYSA
LLFTISGFANQAVIAAMAKEDRIADRLSHASILEASISPSQLRREAHNDYTHAR
LLASPRQGVAVTEGYFNSHGDSPALAEIDQYTOQINGMLMDADAGTYVIEQGR
SCMLQKAPPELLVTFGFGVSGAAVLCSTVADYLLQEPARHLITSTMPVQOAL
RASLVAIRSDGDAKREKLALITFRFAGVODLEPTLADSCSAIOLPLVDNSRAIOL
AEKLRQGCWVTAIRPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNC"
3030..3045
/standard_name="BIOB RBS"
gene
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/translation=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEYGTDTVEGKTVASCALLQAKAAGTYTAGKYVYASG
KTPBGLNUSDALALQKNSQLQDYATVNPPTAEPSPHITSQEGRPLESIVMSAG
LRALEQADWLVLEAGAGWEPPLSDTFADWVYQEOUPVILVGVGLGGINHAMLTA
QVIOHAGITLAGWANDVTPPGKRHAHEMTTLTRMIPAPLGLIEIPMLAENDEVAATGK
YINLAIFYDSTLGTFSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gttcgtgacatcgtctgttgaagcg 25
|||||
Db 173 gttcgtgacatcgtctgttgaagcg 197

```

```

RESULT 10
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
ACCESSION A93679.1 GI:6741867
VERSION A93679.1
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
Escherichia.

```

```

RESULT 11
AR101809 5872 bp DNA PAT 14-FEB-2001
LOCUS Sequence 1 from patent us 6083712.
ACCESSION AR101809
VERSION AR101809.1 GI:12812607
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.

```

```

unclassified.
REFERENCE      1 (bases 1 to 5872)
AUTHORS        Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE          Biotechnological method of producing biotin
JOURNAL        Patent: US 6083712-A 1 04-Jul-2000;
FEATURES       Location/Qualifiers
               source          1..5872
               /organism="unknown"
BASE COUNT     1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match    100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtgtgcatctgctgttgaagc 25
    |||||
Db 173 GTTGTGATCTGCTGTTGAAGC 197

RESULT 12
ARI01810
LOCUS          ARI01810 5872 bp DNA PAT 14-FEB-2001
DEFINITION     Sequence 6 from patent US 6083712.
ACCESSION      ARI01810
VERSION        ARI01810.1 GI:12812608
KEYWORDS
SOURCE         Unknown.
ORGANISM       Unclassified.
REFERENCE      1 (bases 1 to 5872)
AUTHORS        Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE          Biotechnological method of producing biotin
JOURNAL        Patent: US 6083712-A 6 04-Jul-2000;
FEATURES       Location/Qualifiers
               source          1..5872
               /organism="unknown"
BASE COUNT     1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match    100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtgtgcatctgctgttgaagc 25
    |||||
Db 173 GTTGTGATCTGCTGTTGAAGC 197

RESULT 13
AE000180
LOCUS          AE000180 11022 bp DNA BCT 01-DEC-2000
DEFINITION     Escherichia coli K12 MG1655 section 70 of 400 of the complete
               genome.
ACCESSION      AE000180 U00096
VERSION        AE000180.1 GI:1786988
KEYWORDS
SOURCE         Escherichia coli K12.
ORGANISM       Escherichia coli K12.
               Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
               Escherichia.
REFERENCE      1 (bases 1 to 11022)
AUTHORS        Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
               Riley, M., Collado-Vides, J., Glasner, J.D., Koe, C.K., Mayhew, G.F.,
               Gregor, J., Davis, N.W., Klinkpatrick, H.A., Goeden, M.A., Rose, D.J.,
               Mau, B. and Shao, Y.
TITLE          The complete genome sequence of Escherichia coli K-12
JOURNAL        Science 277 (5331), 1453-1474 (1997)
MEDLINE        97426517
PUBMED         9278503
REFERENCE      2 (bases 1 to 11022)

```

```

AUTHORS        Blattner, F.R.
TITLE          Direct Submission
JOURNAL        Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
               University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
               Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
               608-263-7459
REFERENCE      3 (bases 1 to 11022)
AUTHORS        Blattner, F.R.
TITLE          Direct Submission
JOURNAL        Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
               University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
               Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
               608-263-7459
REFERENCE      4 (bases 1 to 11022)
AUTHORS        Plunkett, G. III.
TITLE          Direct Submission
JOURNAL        Submitted (13-OCT-1998) Laboratory of Genetics, University of
               Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
               This sequence was determined by the E. coli Genome Project at the
               University of Wisconsin-Madison (Frederick R. Blattner, director).
               Supported by NIH grants HG00301 and HG01428 (from the Human Genome
               Project and NCHGR). The entire sequence was independently
               determined from E. coli K12 strain MG1655. Predicted open reading
               frames were determined using Genemark software, kindly supplied by
               Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
               30332 (e-mail: mark@ember.gatech.edu). Open reading frames that
               have been correlated with genetic loci are being annotated with CG
               Site Nos., unique ID nos. for the genes in the E. coli Genetic
               Stock Center (CGSC) database at Yale University, kindly supplied by
               Mary Berlyn. A public version of the database is accessible
               (http://cgsc.biology.yale.edu). Annotation of the genome is an
               ongoing task whose goal is to make the genome sequence more useful
               by correlating it with other data. Comments to the authors are
               appreciated. Updated information will be available at the E. coli
               Genome Project's World Wide Web site
               (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
               its annotations are periodically updated; this is version M4. No
               sequence changes. Annotation updates: updated gene identifications
               and products; all new functional assignments courtesy of Monica
               Riley; added promoters, protein binding sites, and repeated
               sequences described in reference 1. The unique numeric identifiers
               beginning with a lowercase 'b' assigned to each gene (protein- or
               RNA-encoding) are now designated as gene synonyms instead of
               labels. This should allow them to be searched for in Entrez as gene
               names.
FEATURES       Location/Qualifiers
               source          1..11022
               /organism="Escherichia coli K12"
               /strain="K12"
               /sub_strain="MG1655"
               /db_xref="taxon:83333"
               <1..130
               repeat_region
               gene
               CDS
               /note="REP (repetitive extragenic palindromic) element;
               contains 4 REP sequences"
               complement(147..1430)
               /gene="ybhC"
               /note="b0772"
               complement(147..1430)
               /gene="ybhC"
               /function="orf. Not classified"
               /note="1427; 96 pct identical to fragment YHC_ECOLI
               SW:P46130 (300 aa) but contains 127 additional C-terminal
               residues"
               /codon_start=1
               /transl_table=11
               /product="putative peptidesterase"
               /protein_id="AAC73859.1"
               /db_xref="GI:1786989"
               /translation="MNPFSVSRLLALAFGVTLNACSTPPDPSPDOTAGTSSRPI
               LSAEADNDAOHYPASLLPGAAWNPSPITLPPQDPVVPAGVGQVTHTTIOAAD
               AAIKTRNKROYIAVMPGEYOCGYVPAAGGILYLGXGEPIDIKVIGLSIDGMSPA
               DMRHIVNPRGKTMPEGPAWMTYRDCOSKRSISIGVLSAVYWSGNNGIOLQNLITEN
               LGDSVDAGNHPVALRTDQVQINNVTIGRQNTFVTNSGVQNRLETNRQPTLVN

```

NSYIEGDVIVSGAVVFDNTEFRVNSRTQEAAYEPAPATLSNIYYGFLAVNSRFN
 AGDGGVAGLRSLDVAANTNGOVIRDSAINCEFNIAKPMADAVISNRPFGAGTGSVD
 DNDEIQRNINNTNRMEXNNRGVSKVAAAKK"

promoter
 complement(1507..1535)
 /note="factor Sigma70: predicted +1 start at 806574"

gene
 complement(1582..2058)
 /gene="ybhB"
 /note="b0773"
 complement(1582..2058)
 /gene="ybhB"

CDS
 /function="orf: unknown"
 /note="1158: 99 pct identical to YBH-B_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"

protein_bind
 /translation="MKLISNDLRDGDKLPRHVPNGMGYDGNISPHLAWDVPAGTK
 SFVYTCYDPAITSGMHWVYVNLPAIDRVLPQSGSLVAMPDGVLOTTRDFGKTG
 YDGAAPKGETHYIFTHALDIERIDVEGASGAWCFNVHFHSLASITAMS"

gene
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rhas predicted site"
 complement(2117..3406)
 /gene="bioA"

CDS
 /note="b0774"
 complement(2117..3406)
 /gene="bioA"

gene
 /note="b0775"
 complement(2117..3406)
 /gene="bioA"

protein_bind
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="1429: 100 pct identical to BIOA_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"

promoter
 /translation="MTTDLAFDQRHIMHPTSMPLPVYVSAEGCELLSDGR
 LVDGSSMAAIGYNNPOLNAKSOIDMSHVMGGITHAALICRKLAMTPOP
 LECVPLADSGVAVYAKMAIOYNKAGRQRLTPFNGYDGFEGDPPNS
 MSLKGYLDPENLFAFPOSMDGENDERMVGFALMAHREIAAVIIEPVQAG
 GMRHYHEWLRIRKICDREGILLIADETATGFRGKLPACRHAETADPILGLGAL
 TGCTMTLSATLTTRVAAETISNGEAGCFMHPFPMGNPLACAAHSLAI1LEGDMQO
 OVADIEVQLREQLAPARDAEMVADVILGIVETTHVNMALQKFEQGVWIRP
 EKLTIEMPYIILLPOOLRLTAANRAVQDDEFQO"

promoter
 complement(2193..2221)
 /gene="bioA"

protein_bind
 /note="factor Sigma70: predicted +1 start at 807260"
 3411..3450
 /note="central position to bioB promoter: -20"

protein_bind
 /bound_moiety="BioB documented site"
 3411..3450
 /note="central position to predicted promoter: -20; genetic
 evidence for the site"

protein_bind
 /bound_moiety="BioB predicted site"
 complement(3411..3450)
 /note="central position to predicted promoter:50"

protein_bind
 /bound_moiety="BioB predicted site"
 complement(3411..3450)
 /note="central position to bioA promoter:50"

promoter
 /bound_moiety="BioB documented site"
 3413..3441
 /note="factor Sigma70: promoter bioB; documented +1 at
 808525"

promoter
 complement(3447..3473)
 /note="factor Sigma70: promoter bioA; documented +1 at
 808515"

gene
 3493..4533
 /gene="bioB"
 /note="b0775"

CDS
 3493..4533
 /gene="bioB"
 /EC_number="2.8.1.-"

/function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0346: 99 pct identical to BIOB_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="Biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"

gene
 /translation="MAHRRPWTLSQVTELEFEKPLDLLPEAQOVHROHPDPOVST
 LLSIKTGACPECKYCPQSSRYKTGLAEERLMEVEVLESARAKAGSTRFCGAAM
 KNHPRDMRYVEGVGVKAMGLEACMTIGTISFSGOARLNGDYNNLDPSPER
 YGNITTTTVEERLDTEKRDAGIKKCSGIVGLGTVDRAGLLLQLNLTPPES
 VPIMLVKVKSTPLADDNDVADFIFITIAVARIMPTSTYRUSAGBNBQTOAC
 PMAGENSIFYGCKLLTPNPEEDKDLQFRKLGNIPQOATVLAGDNDQDRLDQALMT
 PTFDEYNAAL"

CDS
 4530..5684
 /gene="bioF"
 /note="b0776"
 4530..5684
 /gene="bioF"

gene
 /EC_number="2.3.1.47"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0384: 100 pct identical to BIOF_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"

translation="MSWQEKINALDARRAADLRPRYPVAGAGRLVADDDRYLNF
 SSNDLGLSHHPQIIIRAWOGAEQFGTSGSVHGVSVHQALEELAEWLYSAR
 LLEISGFANQAVIAAMMAKEDRIADRLSASLEASLSLSQLRFAHNDVYTHAR
 LIASPCQGOVTEGVFSMDGDSAPLAIEIOVQOIHGMVMDADAGTVIGEOGSG
 SWLQKVKPELLVTFKGFVSGAALVCSSTVADYLLQFARHLIYSTMPPAQOAL

Query Match 100.0%; Score 25; DB 1; Length 1102;
 Best local Similarity 100.0%; Pred. No. 0.0013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgcgagtcagctgttgaagc 25
 Db 3549 GTTGGTGCATCTGCTTTGAAACG 3573

RESULT 14
 LOCUS AE005258
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 of 155.
 ACCESSION AE005258 AE005174
 VERSION AE005258.1 GI:12513751
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 1 (bases 1 to 13501)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A., Blattner,F.R.

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 2 (bases 1 to 13501)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A., Blattner,F.R.

Genome sequence of enterohemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)
 21074935
 11206551
 2 (bases 1 to 13501)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A., Blattner,F.R.

TITLE Welch, R.A. and Blattner, F.R.
 JOURNAL Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetelics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 FEATURES Location/Qualifiers
 source 1..13501

misc_feature

/organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohemorrhagic"
 <1..7576
 /note="O-island #36: Region of the EDL933 chromosome not
 homologous to E. coli K-12 MG1655; Cryptic prophage
 CP-933K; Includes one copy of the 13 bp direct repeat that
 flanks the prophage"
 66..665
 /gene="1omk"
 /note="20981"
 66..665
 /gene="1omk"
 /function="putative membrane; other or unknown (Phage or
 prophage Related)"
 /note="Residues 1 to 199 of 199 are 72.86 pct identical to
 residues 1 to 199 of 199 from Genpept 118 :
 g117532789|gb|AF63231.1|AF151091_2 (AF151091) 1om
 [prophage P-Elba]"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein of prophage
 CP-933K"
 /protein_id="AAG55137.1"
 /db_xref="GI:12513752"

CDS

/translation="MKRYCAIIISATICLAVSGVPWASEHQSITLSTGTLHASTDAPG
 SDDLINGINRYRFTDITGLITFSYANAEDEQKTHYSDTRHEDYVNRWFSVAC
 PSYRVNWFESAYMAGYASRVSTFSGDYFRVTDNKRKTHDVLTGSDARYSNTSLAW
 GAGVQFNPTEFSAVDVAYEXSGSGDWMTDGFIVGVYKF"
 719..2041
 /gene="20982"
 719..2041
 /gene="20982"
 /function="putative structure; Structural component (Phage
 or prophage Related)"
 /note="Residues 164 to 440 of 440 are 68.79 pct identical
 to residues 381 to 645 of 645 from Genpept 118 :
 g114585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
 tail fiber protein [Bacteriophage 933w]"
 /codon_start=1
 /transl_table=11
 /product="putative tail component of prophage CP-933K"
 /protein_id="AAG55138.1"
 /db_xref="GI:12513753"

gene

CDS

/translation="MGMAVOTISGVLCDCAGKPIQCTQLKARNSTVYVNTVASE
 NPPBAGRTSMDEVGYOTVTLLEGGFPSPAGTISYEDSOPTLNDIFGAMTEDAR
 PHLRRFEQWVEEARAEAKNAGEAETISANAGISAKAEASANADTSAEDAS
 SAOAAESASAKSEASASSSSAEOAKAESLSQATDAELSKTAEASAAANADA
 TTEKARESAEASOASQSRILAEDAVNRIPTVYGPCKGPGGPGAGQCGKAGE
 RGDTPAGATGEGRGCDTGGAPDGGPKDGRGRTGLTGACGPGCGPGDGAAGA
 GPGKPGTGAAGATGATGPGKDPGPGTQIRLGPWRILLETNSYGFPTDQALI
 TGLTFLDPKATVQGMFOHLVRFEDGCDWQDYKGLDEGSDTGRGE"
 2004..2312
 /gene="20984"
 2004..2312
 /gene="20984"
 /function="orf: Other or unknown (Phage or Prophage
 Related)"
 /note="Residues 1 to 102 of 102 are 98.03 pct identical to
 residues 28 to 129 of 129 from Genpept 118 :
 g114585437|gb|AAD25465.1|AF125520_60 (AF125520)
 hypothetical protein [Bacteriophage 933w]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"

CDS

gene

gene

CDS

2489..3469
 /gene="20985"
 2489..3469
 /gene="20985"
 /function="orf: Other or unknown (Phage or Prophage
 Related)"
 /note="Residues 13 to 326 of 326 are 55.73 pct identical
 to residues 18 to 331 of 336 from Genpept 118 :
 g116960367|gb|AF33527.1 (AF170116) hypothetical protein
 predicted by Glimmer [Salmonella typhimurium Ur2]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AAG55140.1"
 /db_xref="GI:12513755"
 /translation="MISPIRTTFNSVNIYQSSPCQVYFAGKEKELVIDEPTILF
 QMFEPNPERKKDEVPVNTKOHPIYDVTVAARIESDRMTGIFVDDGFSYKOTAPS
 KLEPDEENMIIREDVDPSMVDRLSDIYHDIICQRLREDEKRDDELNTLEKRL
 ELSKADSLISYAKKRHHAFDFPRNALLAKAGTFRCTYNTNRHGISREGCTYID
 ADMITLTKLITTYIADGISMVDRNDVNTSANSIIVNSNHPALLEGLSPHRSKD
 AHRYDGLCKGVKRYFNFTPLNHNHCDPTEFNHPTIMNTSYTCSSW"
 3530..4522
 /gene="20986"
 3530..4522
 /gene="20986"
 /function="orf: Other or unknown (Phage or Prophage
 Related)"
 /note="No significant matches"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AAG55141.1"
 /db_xref="GI:12513756"

gene

CDS

/translation="MKIPSLDSNFSAPAGYAPAPNRAENYADYVDIGRIPL
 SAADSNVESVIRAVHDSRSRLIDQHTVMIQTVLADLSRSQTFDAYSIGHNK
 VHIGIKRYNEIELNESSVKIDIOSLTLCLELEYEGVDEPIEICEENDEEY
 VSFVAPDPTDSYEMWQEGLIHETIHVTCSPDSGSDNIELGPTTILARRVAEIG
 MSVPDEKGAEPREHRLRLNLAALQAMRHEENRAPERIGCTISDYEASPDPT
 EYASVNTGVGFIOQHDPGLAINDNQDANQIOLYTGAPYITFFGVYDKHNO"
 5350..6231
 /gene="20989"
 5350..6231
 /gene="20989"
 /function="orf: Other or unknown (Phage or Prophage
 Related)"
 /note="No significant matches"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AAG55142.1"
 /db_xref="GI:12513757"

gene

CDS

/translation="MISPSYVNIQGSNLSLRLNITSPDNVLYSSVRAAVHSNGQAY
 KVGNRTIRYVATDKRCVYTRSHSGCTNLHLRIGKPGKETSRTIEVMLASVSAAM
 ERGLVHSRRLPVDVAPLPLPSVDNRLSYSPNVNIGKGNVYEDADARKVAK
 FTTQSNEDEVTEVRCFNOYYGASAKIYGNGDILGIMDRKINGESILNLSILPAQ
 AEHATYMFDRLEQKGLIFVDTETNLVYLRANKFNPIDISSYVNSDRSRSBQIWO
 SYHGKODLISVLSKI"
 6462..7160
 /gene="20990"
 6462..7160
 /gene="20990"
 /function="orf: Other or unknown (Phage or Prophage
 Related)"
 /note="No significant matches"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AAG55143.1"

gene

CDS


```

/db_xref="GI:12513758"
/translation="MRPTSLNVLHSSRSSSDTDIESLWKASVOMIKNPOLRF
OCTDNIAVOIEALDKIGSTETGRVLINATISRLKSEIVLHNSRGVAMHRD
IDAEHNRGSGDFHNLNAVEFPGCEGIVDFHATFVHFHNLNENRKLKVES
SRAESOKSYPLLEAFRTVGLGAFSEEVLENKFEETIGMRRTSYPADSALINDNT
VSLGFOQVRLHPL"
gene      complement(7668..8144)
         /gene="ybhB"
         /note="Z0992"
         /complement(7668..8144)
CDS
         /gene="ybhB"
         /function="orf: Unknown function"
         /note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
         /codon_start=1
         /transl_table=11
         /product="orf, hypothetical protein"
         /protein_id="AAG55144.1"
         /db_xref="GI:12513759"
/translation="MKLISNDLRDGDKLPHRHVFNMGVYDGDNISPHLAMDVPAGTK
STVAVCYDDPADATGSGMMHWVYNNIPADTRVLPQCFSGVLAMPDGVLTQTRDEKTS
YDGAAPKGETHRTLVTHALDVERIDVDEGASGAMGVNHFHSLASAITAFMS"
gene      complement(8203..9492)
         /gene="bioA"
         /note="Z0993"
         /complement(8203..9492)
CDS
         /complement(8203..9492)

Query Match      100.0%; Score 25; DB 1: Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 gttcgtgacatctgtttgaagc 25
        |||||
Db      9635 GTTGCTGAGATCTCTGTTGAAGCG 9659

RESULT 15
LOCUS      AP002553      297816 bp      DNA      BCF      07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION  AP002553 BA000007
VERSION      AP002553.1 GI:13360211
KEYWORDS
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:R1MD 0509952)
           DNA.
ORGANISM    Escherichia coli O157:H7
REFERENCE   1 (sites)
AUTHORS     Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
           Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,T., Abe,H., Iida,T.,
           Yamamoto,C., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
           Sasakawa,C. and Shinagawa,H.
           Complete nucleotide sequence of the prophage VT2-Sakai carrying the
           verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
           derived from the Sakai outbreak
           Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL     20198780
MEDLINE
REFERENCE   2 (sites)
AUTHORS     Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
           Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
           Hayashi,T.
           Comparative analysis of the whole set of rRNA operons between an
           enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
           Escherichia coli K-12 strain MG1655
           Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL     20557356
MEDLINE
REFERENCE   3 (sites)
AUTHORS     Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
           Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
           Yamamoto,C., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
           Shinagawa,H.
           Complete nucleotide sequence of the prophage VT1-Sakai carrying the
           Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
           O157:H7 strain derived from the Sakai outbreak
           Gene 258 (1-2), 127-139 (2000)
JOURNAL     20564182
MEDLINE
REFERENCE   4 (sites)
AUTHORS     Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
           Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
           Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
           Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
           Shinagawa,H.
           Complete genome sequence of enterohemorrhagic Escherichia coli
           O157:H7 and genomic comparison with a laboratory strain K-12
           DNA Res. 8 (1), 11-22 (2001)
JOURNAL     21156231
MEDLINE
REFERENCE   5 (bases 1 to 297816)
AUTHORS     Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
           Hayashi,T.
           Direct Submission
           Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
           Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
           Japan (E-mail:kengen-info.osaka-u.ac.jp,
           URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
           Fax:81-6-6879-2047)
           genome project.
COMMENT     location/Qualifiers
FEATURES
SOURCE      1..297816
           /organism="Escherichia coli O157:H7"
           /strain="O157:H7"
           /sub_strain="R1MD 0509952"
           /db_xref="taxon:83334"
           79..1245
           /gene="ECs0753"
           /note="similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
           /codon_start=1
           /transl_table=11
           /evidence=not_experimental
           /product="succinyl-CoA synthetase beta subunit"
           /protein_id="BAB34176.1"
           /db_xref="GI:13360212"
           /translation="MNLHEYQAKOLFAPRYGLPAPVGYACTTPREAEBAKIGAGPV
VKCOVHAGRGKAGGVVNSKEDIRAFENMLCKRLVYOTANGOPVQILYEAT
DIAREYLVGAVDRSSRRVYVMSSTEGVELEVAERTPHLHKVALDPLTGMPOG
RELAFKIGLBEKLVQFTKIFMGATITFLERDLALITINPLVITTKODLICDGRKGA
DGNALFROPDLREMRDQSDPEPAQAWLENTVALDNGVNGVAGLAWGTMDIV
KLHGEPANFLDVGGATKERYEAFKTIISDKRVKAVLVNIFGIVRCDLADGIIG
AAVEGVAVPVVVRLEGNNALGAKRLADSGLNIIAAKGLTDAAGQVVAVEGK"
           1245..2114
           /gene="ECs0754"
           1245..2114
           /gene="ECs0754"
           /note="similar to SUCC_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
           /codon_start=1
           /transl_table=11
           /evidence=not_experimental
           /product="succinyl-CoA synthetase alpha subunit"
           /protein_id="BAB34177.1"
           /db_xref="GI:13360213"
           /translation="MSLIDKNTKVICGFTGSGTFSRQALAYGRKMGVYTPGKG
GTHHLGLPVFNTVEAAVAAATASVIVAPFCKDSILIEDAGIKLITITIGIPT
LDMILFVYKIDAGVAVMGPCPVITPGCKGICIGKHIKKGKGVIVRSSTIYE
AVKOTTDYGFCSQTCVIGDPIPGSNFTIDLEFEDPOTBEATVWIGELGSAEBA
LAAYIKENHTKPVVYIAGVTAIPKGRMGHAGAITAGSKGTADEKFALEAGVKTAYS
LADIGELAKTVLK"
           2781..3686
           /gene="ECs0755"
           2781..3686
           /gene="ECs0755"
           /note="probable transcriptional regulator, similar to

```

```

transcriptional regulators e.g. glycine cleavage system
transcription activator (qcv operon activator) -
Escherichia coli g14170431sp1P320641GCVA_ECOLI1 percent
identity 31 in 300 aa
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKLPKTELVTFEVARHESYTRAEELATQSAFVROVSAI,
EELRPPLFSHKKRIPLNDYGYVYKVEITNKLEPNTMTWQPIVOYTELVN
PTSTHWLIPNLHEPTKIPDITVINHSANNDDIPREVDATIMENCAPIAEVY
LPEBEILPVCSGLIAMSODKLSVAELTELPLTHOSTRTGHEEPALSGVSSPLVN
NGPFDLSMLIAVANSNIGVALLPREFALQHDSDGDMV1PCDVP1RTGNRPLTMWQ,
EKSDSPHLOQFREMILAKSVYQDM"
/complement(3720..4322)
/gene="ECS0756"
/complement(3720..4322)
/gene="ECS0756"
/gene="ECS0756"
/translation="MEARISTPRHVOROOKIKFOVTRVAALAEKKGLIVFGNKS
KSTAAEGTVRAHGKTVGVAOYIKGMDNGETNLLQPLCYEFHIMGTFWETONK
QADIDAKKEVWSESKRMALADKRDVLVLDLTYMAHYLDVEVIASLONRPAQOSV
IVTGRCHSQLIMADTVSEIRPVKHAFFNGIOAPGIDM"
/complement(4332..5984)
/gene="ECS0757"
/complement(4332..5984)
/gene="ECS0757"
/gene="ECS0757"
/translation="probable fumarate hydratase, similar to
fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g11205981sp1P009231FUMA_ECOLI1 percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPTWQEPFLONKDGPEYTLISDQHTVTELDGEYIKIAP,
ATLLARQAFVASFPLSAHLQOVASILINDOASSNDKYVALQILRNAYEAKGLP
NCDPTATIVASKGOQIRTGNDALSKGITYSTQENNLPSQAPADMTVEVNTQ
TNLPAQIDISAVAGDEYHPLCVNKGGSANKALVQETKSLQPEKILAFLEKKSL
GTACPPYHIAFVGGISADOTLKIAKLASTKYDNLPTSGNDQGAFFDILEVLL
EASQOFGIGAOGKGYFAHDIVYRLPRHGSCLPAMALSCSADRIRAKIKHGHLI
EXLEHNHGOYIPASIRENNAOHVOLDLNRPLRDMODLARLPVGRVSLSGPIYVAR
DIAHAKIKARLOSCEPMPEYIKHHIYVYAGPAKTPENMACSLGPTTGRMGYIDLP
QAAAGSLVHLSKNSQOYTDACHKHGNGLSIGAAALLQETVKSURCLEYBELG
MEAVMVEVNDPAFTLVDDKGNPFQEFQDHKASCAPAH"
/complement(6092..7372)
/gene="ECS0758"
/complement(6092..7372)
/gene="ECS0758"
/translation="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g11214671sp1P249431GLTT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative transport protein"

```

```

Query Match
Best Local Similarity 100.0%: Score 25; DB 1; Length 297816;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgagatcgtcgtttgaagc 25
DB 92367 GTTGCTGATCTGCTGTGGAAGC 92391

RESULT 16
LOCUS E00893 1121 bp DNA PAT 29-SEP-1997
DEFINITION Genomic DNA encoding biotin synthetase.
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirano, Y., Kojima, T., and Kimura, H.
TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
JOURNAL PRODUCTION OF BIOTIN
COMMENT NIPPON SODA CO LTD
OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PI HIROMO YOSHIOHKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12P13/19), (C12P13/18, PC
C12P13/19);
CC strandedness: Double;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=Escherichia coli Ns101;
CC feature is identified by experimental;
FH key Location/Qualifiers
FT CDS 42..1079
FT /product="biotin synthetase".
FEATURES
source Location/Qualifiers
1..1121
/oranism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 289 a 296 c 325 g 211 t
ORIGIN
Query Match 96.0%: Score 24; DB 6; Length 1121;

```

Best Local Similarity 100.0%; Pred. No. 0.0055;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtctgctgcatcgcgtttgaagc 24
|||||
Db 98 gttctgcatcgtctgttgagc 121

RESULT 17

AE004192

LOCUS 12891 bp DNA BCT 31-JUL-2000
DEFINITION Vibrio cholerae chromosome I, section 100 of 251 of the complete
chromosome.
ACCESSION AE004192 AE003852
VERSION AE004192.1 GI:9655581
KEYWORDS
SOURCE Vibrio cholerae.
ORGANISM Vibrio cholerae.
REFERENCE 1 (bases 1 to 12891)
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
AUTHORS Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermoлаeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D.,
Nierman, W.C. and White, O.
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
Nature 406 (6795), 477-483 (2000)

TITLE

JOURNAL MEDLINE
20406833
2 (bases 1 to 12891)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermoлаeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
20406833
2 (bases 1 to 12891)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermoлаeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA

FEATURES

source
Location/Qualifiers
1. 12891
/organism="Vibrio cholerae"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/note="biotype: El Tor"
complement(84. 1370)
/gene="VC1111"
complement(84. 1370)
/gene="VC1111"
/note="similar to GB:J04423 SP:P12995 PID:457106
PID:490217 GB:000096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="adenosylmethionine-8-amino-7-oxononanoate
aminotransferase"
/protein_id="AAF94270.1"
/db_xref="GI:9655582"
/translation="MMMDLAFDRCHIMHYTSTLPPLTCYPSVRADGVIITLEDGRSL
VQMSGMAAIGHYNIHPRKLNQAAHOIEQMSHMGILTHOPATLCOKLLKLANNL
OOVFLADSSVAVESLAKALOYHWSKQCPRAKFLTLRGYTHGDTFAANSYDPPNSM
HSLYKGLPEHIFANSPEGGFDPAMBERDIOFRKHLETHHIOIAVILIEPVOAGG
MRITPEFLROYALCDERGVALLIDELITGGRGKLPFACHEAGIOPDILCVGALT
CGMYLTSATLTQOAVDFVAGSGLCFMIEGTPGNPLACAVASASLSIEGDWQQ
VANIEAFEAOLPKLNDSPRKOTRWLGAIVGVEHLPVNMEVIOALFVEHGWIMRPF
GRITLYMPYISQPEHLKLVDAAMAALNQPCFRQ"

gene

CDS

gene
CDS
/gene="VC1112"
1497. 2549
1497. 2549

gene

CDS

/gene="VC1112"
/note="similar to GB:J04423 SP:P12995 PID:145425
PID:490219 GB:000096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="AAF94271.1"
/db_xref="GI:9655583"
/translation="MEYRNHVAEVKALLDKRPMDLPEAOOVHRLHPRHNVOST
LLSTKTGACPEDCYCPQSAHYRVDVKEKLEMEYRVLDAQAKKSGSTRFCMAAM
KNPERMDPLKEMIRGVKDMGLCTQMLDMIPDAQQAQAGLDLYNNINDTSEF
YGNITITRYQDLRLDLSHVADGAKICSGGIIIGMGESTRNDAGLLVLELITPDES
VPINMLVKVGTPLEQVDVPEPFRILVAVRIMPSAVRLSAGREKNEQOALC
FMAGANSIFYGCKLITTPNPAEDSDMLFFKIGINRQVAKRDEITENLDRIYVER
VAAPRTASDLFYDAL"

gene

CDS

gene
CDS
/gene="VC1113"
2533. 3687
2533. 3687

gene

CDS

gene
CDS
/gene="VC1115"
4463. 5179
4463. 5179

gene

CDS

gene
CDS
/gene="VC1116"
5261. 5470
5261. 5470

gene

CDS

```

/gene="VC1116"
/note="identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AF94275.1"
/db_xref="GI:9655587"
/transl_table="MSRRNTTNKSKHMAEHSYOMAILDLCCHGISDFEARTQUGL
EAKETLQOQIVTQHSIMKLSAEK"
5697..6560
/gene="VC1117"
5697..6560
/gene="VC1117"
/note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
PID:1736470; identified by sequence similarity: putative"
/codon_start=1
/transl_table=11
/product="heat shock protein HtpX"
/protein_id="AF94276.1"
/db_xref="GI:9655588"
/transl_table="MKRILLFLATNLAAYVLVSVINIYAVTGMOPGSLGLYMAA
VEFGGAFISLMSKSMALRSVGGVITDPRNEMHEHLLTVRRQANQAGICMPYAI
YDAPDMNAEFTGKRDSDLVAVSTGLHMMTRDEAVLAHEVSHIANDQMTLMQ
GVNTEVFILSRPIANIVASRDEEGEGSNMVMYFGVSVLETFEGLASFITMYYSR
HREFHADGAAOLGKHKMIALERLKMQESHLESGMMAFGITGKRSLSLEIMTHPP
LEKRIALRRM"
6770..7198
/gene="VC1118"
6770..7198
/gene="VC1118"
/note="similar to PID:1653608; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator. putative"
/protein_id="AF94277.1"
/db_xref="GI:9655589"
/transl_table="MSIMDVQNAOFYQOLDKSQLRLIEIYHPDVFEDAHRIEGF
DALYQELNLYQNVHTCTPTIHEQVAVNEGAFIWMTHLRHPKLAKGEVDVAKGSHL
HFAEGKVTYHRDQFDMGEMLYEQLPYLCGVIRAIKRLGQ"
7195..7920
/gene="VC1119"
7195..7920
/gene="VC1119"
/note="similar to SP:P25970; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, short-chain
dehydrogenase/reductase family"
/protein_id="AF94278.1"
/db_xref="GI:9655590"
/transl_table="MTAVFITGATSGICQALADYAKGQNOVIACGRNSVLDSTFO
YANIFPLAFVDYDHGSTRKAAALAOPLQPELWLLNAGDCYTDGKMDVTLMARVNTN
VLGVAIVETGICPHUSCGHRAIVGSIASELALPRAEAGASKAAVAIYARTLQDMK
PLGIEVITLIFPGVATPLIDRMTFAMPMLITVERAAQELTKAGLVSQLVFPKRFWT
LRLGLALPYAGRLVRRLKA"
7931..9211
/gene="VC1120"
7931..9211
/gene="VC1120"
/note="identified by Glimmer2: putative"
/codon_start=1
/transl_table=11

```

Query Match 80.0%; Score 20; DB 1; Length 12891;
 Best Local Similarity 100.0%; Pred No. 0.68;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ttgattctgtcttgaagc 25
 |||
 DB 1558 TGGATCTGCTGTTGAAGC 1577

```

RESULT 18
S64534/c 383 bp DNA INV 27-APR-2000
LOCUS S64534
DEFINITION Drosophila melanogaster defective chorion-1 eggshell protein
(dec-1) gene, dec-1-Fc4 allele, partial cds.
ACCESSION S64534
VERSION S64534.1 GI:407575
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 383)
AUTHORS Andersson, S. and Lambertsson, A.
TITLE Evolution of the dec-1 eggshell locus in Drosophila. II.
INTRASPECIFIC DNA sequence analysis reveals length mutations in a
repetitive region in D. melanogaster
JOURNAL J. Mol. Evol. 36 (6), 536-544 (1993)
MEDLINE 93353525
PUBMED 8350348
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI gibseq 136269] from the original journal article.
This sequence comes from Fig. 4.
Map location: X:7C3-4
FEATURES
source Location/Qualifiers
1..383
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
<1..>383
/gene="dec-1"
/allele="Fc4"
<1..>383
/gene="dec-1"
/note="This sequence comes from Fig. 5: follicle cell
protein for proper eggshell assembly; conceptual
translation presented here differs from translation in
publication"
/codon_start=1
/product="defective chorion-1 eggshell protein"
/protein_id="AAB27805.2"
/db_xref="GI:7548373"
/transl_table="QNPMMMOOROWSEFEAKIOONQOIOONPMMOOROWSEFEAKI
QONQOIOONPMMOOROWSEFEAKIOCHIDHOMQOQAQCSIMTEBOROROWSEDOAKI
QQAQQAQOQPPMMPOMQOROWTEDP"

```

BASE COUNT 143 a 90 c 102 g 48 t
 ORIGIN

Query Match 68.0%; Score 17; DB 3; Length 383;
 Best Local Similarity 100.0%; Pred No. 38;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgctgattctgtctgt 18
 |||
 DB 159 TTGCTGATCTGCTGTT 143

RESULT 19
 S64532/c 461 bp DNA INV 27-APR-2000
 LOCUS S64532
 DEFINITION Drosophila melanogaster defective chorion-1 eggshell protein
 (dec-1) gene, dec-1-Fc3 allele, partial cds.
 ACCESSION S64532
 VERSION S64532.1 GI:407573
 KEYWORDS fruit fly.
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 461)

[illegible][illegible]

KEYWORDS dec-1 locus; tandem repeat.
 SOURCE Drosophila sechellia.
 ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 774)
 AUTHORS Escher, S.A. and Lambertsson, A.
 TITLE Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence comparisons of the simulans complex repeated domain reveal non-concerted evolution
 JOURNAL Hereditas 124 (2), 111-120 (1996)
 MEDLINE 96376325
 REFERENCE 2 (bases 1 to 774)
 AUTHORS Escher, S.A.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of Genetics, S-901 87 Umea, SWEDEN
 COMMENT Related sequences: M35889, M35688 and X07697.
 FEATURES
 source 1..774
 /organism="Drosophila sechellia"
 /db_xref="taxon:7238"
 /chromosome="X"
 /map="7C"
 /map="7C"
 repeat_region 1..774
 /rpl_type=TANDEM
 gene 1..774
 /gene="dec-1"
 <1..>774
 /gene="dec-1"
 /number=2
 exon
 BASE COUNT 293 a 178 c 213 g 90 t
 ORIGIN
 Query Match 68.0%; Score 17; DB 3; Length 774;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ttgttgatctgtctgt 18
 |||||||
 Db 81 TTGCTGATCTGCTGTT 65
 RESULT 25
 DSDC1TR2/c 852 bp DNA INV 25-SEP-1996
 LOCUS D.simulans DNA from repeated region of dec-1 locus, strain
 DEFINITION Yugoslavia.
 ACCESSION X96927
 VERSION X96927.1 GI:1262760
 KEYWORDS dec-1 locus; tandem repeat.
 SOURCE Drosophila simulans.
 ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 852)
 AUTHORS Escher, S.A. and Lambertsson, A.
 TITLE Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence comparisons of the simulans complex repeated domain reveal non-concerted evolution
 JOURNAL Hereditas 124 (2), 111-120 (1996)
 MEDLINE 96376325
 REFERENCE 2 (bases 1 to 852)
 AUTHORS Escher, S.A.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of Genetics, S-901 87 Umea, SWEDEN
 COMMENT Related sequences: M35889, M35688 and X07697.
 FEATURES
 source 1..852

/organism="Drosophila simulans"
 /strain="Yugoslavia"
 /db_xref="taxon:7240"
 /chromosome="X"
 /map="7C"
 /map="7C"
 repeat_region 1..852
 /rpl_type=TANDEM
 gene 1..852
 /gene="dec-1"
 <1..>852
 /gene="dec-1"
 /number=2
 exon
 BASE COUNT 324 a 195 c 232 g 101 t
 ORIGIN
 Query Match 68.0%; Score 17; DB 3; Length 852;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ttgttgatctgtctgt 18
 |||||||
 Db 237 TTGCTGATCTGCTGTT 221
 RESULT 26
 DSDC1TR3/c 852 bp DNA INV 25-SEP-1996
 LOCUS D.simulans DNA from repeated region of dec-1 locus, strain
 DEFINITION Barcelona.
 ACCESSION X96928
 VERSION X96928.1 GI:1262761
 KEYWORDS dec-1 locus; tandem repeat.
 SOURCE Drosophila simulans.
 ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 852)
 AUTHORS Escher, S.A. and Lambertsson, A.
 TITLE Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence comparisons of the simulans complex repeated domain reveal non-concerted evolution
 JOURNAL Hereditas 124 (2), 111-120 (1996)
 MEDLINE 96376325
 REFERENCE 2 (bases 1 to 852)
 AUTHORS Escher, S.A.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of Genetics, S-901 87 Umea, SWEDEN
 COMMENT Related sequences: M35889, M35688 and X07697.
 FEATURES
 source 1..852
 /organism="Drosophila simulans"
 /strain="Barcelona"
 /db_xref="taxon:7240"
 /chromosome="X"
 /map="7C"
 /map="7C"
 repeat_region 1..852
 /rpl_type=TANDEM
 gene 1..852
 /gene="dec-1"
 <1..>852
 /gene="dec-1"
 /number=2
 exon
 BASE COUNT 324 a 195 c 232 g 101 t
 ORIGIN
 Query Match 68.0%; Score 17; DB 3; Length 852;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

* 5265 55000: contig of 49736 bp in length
* 55001 55100: gap of unknown length
* 55101 84235: contig of 29135 bp in length
* 84236 84335: gap of unknown length
* 84336 104093: contig of 19758 bp in length
* 104094 104193: gap of unknown length
* 104194 117951: contig of 13758 bp in length
* 117952 118051: gap of unknown length
* 118052 126732: contig of 8681 bp in length
* 126733 126832: gap of unknown length
* 126833 135507: contig of 8675 bp in length
* 135508 135607: gap of unknown length
* 135608 141357: contig of 5750 bp in length
* 141358 141457: gap of unknown length
* 141458 145691: contig of 4234 bp in length
* 145692 145791: gap of unknown length
* 145792 149740: contig of 3948 bp in length
* 149740 149840: gap of unknown length
* 149840 152725: contig of 2886 bp in length
* 152726 152825: gap of unknown length
* 152826 155298: contig of 2473 bp in length
* 155299 155398: gap of unknown length
* 155399 158289: contig of 2891 bp in length
* 158290 160296: gap of unknown length
* 160297 160396: contig of 1907 bp in length
* 160397 161540: gap of unknown length
* 161540: contig of 1144 bp in length.

```

```

FEATURES
  source
    1..161540
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="3"
      /clone="RP11-220014"

```

```

BASE COUNT      53135 a 28040 c 27770 g 51176 t 1419 others
ORIGIN

```

```

Query Match      68.0% Score 17: DB 2: Length 161540;
Best Local Similarity 100.0% Pred. No. 24;
Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 tgcctgacatcgtcgtt 19
        |||
Db      22839 TGCTGATCTGCTGTTT 22823

```

```

RESULT 33
AE003442      306267 bp      DNA      INV      04-OCT-2000
LOCUS      Drosophila melanogaster genomic scaffold 142000013386054 section 26
DEFINITION      of 35 complete sequence.
ACCESSION      AE003442 AE0025566
VERSION      AE003442.1 GI:7290831
KEYWORDS      HTG.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 306267)
  Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
  Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
  George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
  Sutton,G.G., Wortman,J.R., Vandal,M.D., Zhang,Q., Chen,L.X.,
  Brandon,R.C., Rogers,Y.H., Blazee,R.C., Champe,M., Pfeiffer,B.D.,
  Wan,K.H., Doyle,C., Baxter,E.G., Heit,G., Nelson,C.R., Gaber
  Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
  Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
  Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
  Botchan,M.R., Bouck,J., Brooksstein,P., Brothier,P., Burks,K.C.,
  Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
  Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de

```

```

Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Foster,C., Gabrielian,A.E., Gary,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwan,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kenshion,D.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Krautz,S., Kulp,D., Lai,Z., Lasto,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacib,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Klamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svitskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,C.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveril,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M., and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 306267)
REFERENCE
  Adams,M.D., Celisner,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  Direct Submission
  Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
  Rockville, MD, USA

```

```

FEATURES
  source
    Location/Qualifiers
      1..306267
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /chromosome="X"
        join(32686..32823,32937..33134,33225..33387,34597..34762,
        35618..35749)
        /gene="CG1402"
        /note="Partial transcript."
        /product="CG1402"
        /db_xref="FLYBASE:FBan0001402"
        /db_xref="FLYBASE:FBgn0029962"
        /db_xref="FLYBASE:FBgn0029962"
        <32686..>35749
        /gene="CG1402"
        /map="7C3-7C4"
        /db_xref="FLYBASE:FBan0001402"
        /db_xref="FLYBASE:FBgn0029962"
        join(33264..33387,34597..34762,35618..35639)
        /note="CG1402"
        /gene="CG1402"
        /codon_start=1
        /db_xref="FLYBASE:FBan0001402"
        /db_xref="FLYBASE:FBgn0029962"
        /protein_id="AAF46274.1"
        /db_xref="GI:7290832"
        /translation="MSENRKSGITVIGSLWQIGETPNPELITTSFNKVLNGFS
        TPIRHSVSLPNTDHTTTCGTTTHGCESTWIIYNKPIITTKOEVSLCTCNSD
        D"
        complement(48924..>50117)
        /gene="CG10920"
        /product="CG10920"
        /db_xref="FLYBASE:FBan0010920"
        /db_xref="FLYBASE:FBgn0029963"
        complement(48924..>50117)
        /gene="CG10920"
        /map="7C3-7C5"
        /db_xref="FLYBASE:FBan0010920"
        /db_xref="FLYBASE:FBgn0029963"
        complement(48924..50033)
        /gene="CG10920"

```

mrna

gene

CDS

mrna

gene

CDS

```

gene
    /db_xref="FlyBASE:Fpan0002175"
    /db_xref="FlyBASE:FBgn00000427"
    complement(<63233..>69146)
    /gene="dec-1"
    /note="CG2175"
    /map="7C4-7C5"
    /db_xref="FlyBASE:Fpan0002175"
    /db_xref="FlyBASE:FBgn00000427"
    complement(join(63233..63902,64113..65786,65862..65986,
        66185..66447,66505..68854))
    /gene="dec-1"
    /note="dec-1 gene product; Nucleotide sequence of the
        Cerebra sequence differs from the published sequence for
        this transcript"
    /codon_start=1
    /db_xref="FlyBASE:Fpan0002175"
    /db_xref="FlyBASE:FBgn00000427"
    /protein_id="AAF6278.1"
    /db_xref="GI:7290836"
    /translation="MMPFALGSLIKMGSLFPAYSIIGATIPDNLOPTAAASGVVLYLAD
        DAVKATVARONPNPPLGLMLMPPALPOFOPLPSMDLGPVCSSFLAOLDPAMPVPELT
        LGAAPVPAPAPAAPAPPAPADPPAPYPADQPAIILGOALONAFTFFNPAN
        FDSLSLGQSVTFAPENLDVFVOMQGFEPGGTAPDPAAGTDQAASDISVEVRRE
        DPVSQEOKMKTSALEMEQRQQOAKYKDEOVPLLFRNPTTONDATEEKLIEDLR
        VEKLAIFEROVIATLELMQKIELMAKOMSSAAANGSDSPYSISLPSTPIHKITR
        ADIEQLRDIVYRLVLVKFEARPARNSGIWTKANALIKRAK SODOTISKEDIIVOJIA
        YARMANOMESEKKODKYVAVRTQONPMOROWSEEOAKIQONOOIQONPMMOOROW
        MOOROWSEEOAKIQONOOIQONPMVOCROWSEEOAKIQONOOIQONPMMOOROWSEEOAK
        SEEDAKIQONOOIQONPMVOCROWSEEOAKIQONOOIQONPMMOOROWSEEOAK

Query Match      68.0%; Score 17; DB 3; Length 306267;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 ttgctgatcgcgcgtt 18
          |||||||
Db       67205 TTGCTGATCTGCCTT 67221

RESULT_34
LOCUS      DMDEC1TR5     774 bp      DNA              INV            25-SEP-1996
DEFINITION D.mauritiana DNA from repeated region of dec-1 locus.
ACCESSION X96930
VERSION    X96930.1 GI:1262759
KEYWORDS   dec-1 locus; tandem repeat.
SOURCE     Drosophila mauritiana.
ORGANISM   Drosophila mauritiana
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephyridioidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 774)
            Escher,S.A. and Lambertsson,A.
            Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
            Hereditas 124 (2), 111-120 (1996)
JOURNAL    Journal
MEDLINE    96376325
REFERENCE  2 (bases 1 to 774)
            Escher,S.A.
            Direct Submission
TITLE      Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
JOURNAL    Genetics, S-901 87 Umea, SWEDEN
COMMENT    Related sequences: M35889, M35688 and X07697.
FEATURES
source
    1..774
    /organism="Drosophila mauritiana"
    /db_xref="taxon:7226"
    /chromosome="X"
    /map="7C"
    1..774
    /rptL_type="TANDEM
    1..774

```

exon /gene="dec-1"
<1 >774
/gene="dec-1"
/number=2

BASE COUNT 290 a 182 c 212 g 90 t
ORIGIN

Query Match 64.0%; Score 16; DB 3; Length 774;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ttctgctgttgaagcg 17
|||||
Db 159 TTGCTGTGATCTGCTGT 144

RESULT 35

AF250347/c 1044 bp mRNA VRT 01-JUL-2001

LOCUS Xenopus laevis transcription factor Nkx-2.1 mRNA, complete cds.

ACCESSION AF250347.1 GI:14578835

VERSION

KEYWORDS

SOURCE

ORGANISM

AFRICAN clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 1044)

Hollmann, T. and Pieler, T.

Nkx-2.1: a homeobox gene expressed during early forebrain, lung

and thyroid development in Xenopus laevis

Dev. Genes Evol. 210 (11), 579-581 (2000)

2110556

11180810

2 (bases 1 to 1044)

Hollmann, T. and Pieler, T.

Direct Submission

Submitted (29-MAR-2000) Entwicklungsbiologie, University of

Goettingen, Humboldtallee 23, Goettingen, Lower Saxony 37073,

Germany

Location/Qualifiers

1..1044

/organism="Xenopus laevis"

/db_xref="taxon:8335"

1..29

/note="Nkx-2.1"

1..1044

/codon_start=1

/product="transcription factor Nkx-2.1"

/protein_id="AAK69025.1"

/translation="MSMSPKHTTTPFSVSDILSPLEESYKKVAMGAGLAPLTAAYR

OSOVSAQSMQOHGNGHNGPVSAAVHMTAGVPQUSHTTGWGYCNGNGLGMSGLP

YODTRNSATGNGVAGNDPFRSTISRMGSGGNGMGLGMSGLGDKGSMPLQA

TPRRKRIVLESOAVYELERPRKQKYLSEPEREHLASMIHLPTQVKTIPQNRKRM

KROAKDKASQOQOQDSSCCOQOSPRVAVPVLVKDGKPCQAGSNPTPALQSHQOQ

TATATTTTNGIGPHOSHQTNAGSGSPDLVPHNSPSSLQNVYSLSHLNSSSSDYGS

AMSCSTLYGRTW"

BASE COUNT 277 a 311 c 265 g 190 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 5; Length 1044;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 tctgctgttgaagcg 25

|||||

Db 556 TCTGCTGTGAGCG 541

RESULT 36

AF281080/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 1614)

Small, E.M., Vokes, S.A., Garrick, R.J., Li, D. and Krieg, P.A.

Developmental expression of the xenopus Nkx2-1 and Nkx2-4 genes

Mech. Dev. 96 (2), 259-262 (2000)

20417824

2 (bases 1 to 1614)

Small, E.M., Vokes, S.A. and Krieg, P.A.

Direct Submission

Submitted (21-JUN-2000) Cell Biology and Anatomy, University of

Arizona Health Sciences Center, 1501 N. Campbell Ave., Tucson, AZ

85724, USA

Location/Qualifiers

1..1614

/organism="Xenopus laevis"

/db_xref="taxon:8335"

/note="expressed in ventral forebrain, lungbud primordia

and thyroid anlage"

1..1614

/gene="Nkx2-1"

162..1205

/note="Nkx2-1"

/note="thyroid transcription factor; TTF-1; T/ebp; chicken

and mouse Nkx2-1 ortholog; contains homeodomain"

/codon_start=1

/product="homeobox transcription factor Nkx2-1"

/protein_id="AAK17405.1"

/db_xref="GI:10442650"

/translation="MSMSPKHTTTPFSVSDILSPLEESYKKVAMGAGLAPLTAAYR

OSOVSAQSMQOHGNGHNGPVSAAVHMTAGVPQUSHTTGWGYCNGNGLGMSGLP

YODTRNSATGNGVAGNDPFRSTISRMGSGGNGMGLGMSGLGDKGSMPLQA

TPRRKRIVLESOAVYELERPRKQKYLSEPEREHLASMIHLPTQVKTIPQNRKRM

KROAKDKASQOQOQDSSCCOQOSPRVAVPVLVKDGKPCQAGSNPTPALQSHQOQ

TATATTTTNGIGPHOSHQTNAGSGSPDLVPHNSPSSLQNVYSLSHLNSSSSDYGS

AMSCSTLYGRTW"

BASE COUNT 475 a 407 c 399 g 333 t

ORIGIN

Query Match 64.0%; Score 16; DB 5; Length 1614;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 tctgctgttgaagcg 25

|||||

Db 717 TCTGCTGTGAGCG 702

|||||

RESULT 37

AR120145

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Unknown.

Unclassified.

1 (bases 1 to 5000)

Richards, E.J. and Jeddeloh, J.A.

RNA methylation gene from plants

Sequence 1 from patent US 6153741.

AR120145

AR120145

AR120145.1 GI:14102844

16-MAY-2001


```

/protein_id="CAB56484.1"
/db_xref="GI:5921566"
/translation="MSDNTTKPKPIPKPIRTLRPLVACAEIIRANOKDHPESYLE
OHILVTLQKWKGRFTHQIKKELETASKPVLGLCTLVGSKTLIGEETINMTIRRT
ALPGVARRRGYSVNTLFPYLPVRYMGKILAKIMREYPHLVEYDEDEPVSPYWKER
VIKTFPVKEDKFTALEGTAJHLAIFYVGSYOLSKRMRYVFGHRLDKNEPRIG
YEMGLIEFARFATSEVOTGREYALGLEKSEKGEKEDEKAVVPPKRSSTPIE
DEGETEDKIDLEDPPOLKPIPEASRACITCLSYISAPACTPCGHFCMDCISEWRE
KPECPLCRGVKRONLLPIR"
BASE COUNT      1793 a      1841 c      1485 g      1768 t
ORIGIN

Query Match      64.0%  Score 16;  DB 8;  Length 6887;
Best Local Similarity 100.0%;  Pred. No. 1.1e+02;
Matches 16;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Oy      8  gatcgcctgttgaag 23
|||||
Db      4051  GATCTGCTGTTGAAG 4066

RESULT  40
AE004322  11328 bp  DNA  BCT  31-JUL-2000
LOCUS     Vibrio cholerae chromosome I, section 230 of the complete
DEFINITION
ACCESSION AE004322 AE003852
VERSION    AE004322.1 GI:9657104
KEYWORDS
SOURCE     Vibrio cholerae.
ORGANISM   Vibrio cholerae.
REFERENCE  1 (bases 1 to 11328)
AUTHORS    Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoli,I.,
Sellers,P., McDonald,L., Utterback,T., Fleischmann,R.D.,
Nielsen,W.C. and White,O.
TITLE      DNA sequence of both chromosomes of the cholera pathogen Vibrio
           cholerae
JOURNAL    Nature 406 (6795), 477-483 (2000)
MEDLINE    20406833
REFERENCE  2 (bases 1 to 11328)
AUTHORS    Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoli,I.,
Sellers,P., McDonald,L., Utterback,T., Fleischmann,R.D.,
Nielsen,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
TITLE      Direct Submission
JOURNAL    Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
           Medical Center Dr, Rockville, MD 20850, USA
FEATURES
   source          location/qualifiers
                   1..11328
                   /organism="Vibrio cholerae"
                   /strain="N16961"
                   /serotype="O1"
                   /db_xref="taxon:666"
                   /chromosome="I"
                   /note="biotype: El Tor"
                   complement(95..286)
                   /gene="VC2521"
                   complement(95..286)
                   /gene="VC2521"
                   /note="identified by Glimmer2; putative"
                   /codon_start=1
                   /transl_table=11
                   /product="hypothetical protein"
                   /protein_id="AAF95663.1"
                   /db_xref="GI:9657105"

gene
CDS
/translation="MAANFDVLGIKREVEVINYPPYLAIPALVPCHSKROPLKIVRLS
SACYRLCMKKPFIQFLDAK"
/gene="VC2522"
524..1522
/gene="VC2522"
524..1522
/gene="VC2522"
/codon_start=1
/note="identified by Glimmer2; putative"
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95664.1"
/db_xref="GI:9657106"
/translation="MAFIFYVGINLIEAVVLEFGLAILVWGADKLVEGSAALAR
NIGISPIVAVMTILAMGSSAPEMVYSATALEGKTPDPAVGSNIAINILIGITA
IVKPLVSGAVIRRELPMWIGVTLIAGAIIIDNHNHGEYEGILITLAFAPILMLOY
REKONGDALDEQSEIIPVGSYNKAPRWVYGLILPICAGMIDNAVYIARYFGM
SDLVGLITIAVGSILPELALSLAGALGEDDMAVNTIGSNVFNILAVKGLPILRP
SLSPFAMSRDPMVLMGVSLILVVALGKKRQIRLBGALLCSFTIGITYLLMNI"
1572..2552
/gene="VC2523"
1572..2552
/gene="VC2523"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95665.1"
/db_xref="GI:9657107"
/translation="MRPSPFDYQTVAKOVLATELHAILOOEYIINDPAPACAMILA
NOTGRVVMGKSGHIGKRIATLASTGTSAPFVHPGASHGDLGMEERDIIAITS
NSGESEILALPVLRUSIRVISMTPGNNSMAKLADIHLQITVPACPELAPTS
STATILVMDGDAVALMOARCFEADFDALSHPGALGKRLIKLINDHSGDALPKVA
POLLRDALLETISOKGIMTAIVDEODPLTIGFTFGDRLIDKRDIHSYIVAVMT
RQPTVAQPHLAVEGILNMAKRIINGLMVFNKKLVGALNMHDLKAGVM"
2352..3109
/gene="VC2524"
2352..3109
/gene="VC2524"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95666.1"
/db_xref="GI:9657108"
/translation="MSTVSTLYGVEPSLIEIAKOIKLLTODVGVESDGLIYMGNO
GELKTFHTRDGYKALMNACIETAITGRSQSYENRMRKLGSLIYQGGDDPVQA
YVDICOKLAIADPQYCTIGDLDLMPVMEKVALRVCAVDGHPLLAQRANVYTHIKGH
GAVREYCDLILQARNELDVHKGISL"
3106..3669
/gene="VC2525"
3106..3669
/gene="VC2525"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95667.1"
/db_xref="GI:9657109"
/translation="MSLSRIYVYLLFIASNSLYIILGEOQSKIQVAPNLELPFSG
ENLENTSYDEQIGIRNVITSIHLDHYAKGNTLFPAPILKYRSTGLDWEYIARGI
LSKDQVLTLYDDVLAKNLPLDSGFDPLTYSSENSIOLKSRDFVADRFLRGPFETHG
QAMKGNFADHSAELYNQVGRYETLTP"
3605..4144
/gene="VC2527"
3605..4144
/gene="VC2527"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95668.1"
/db_xref="GI:9657110"
/translation="MPIIRLNFTYKXVDMKRSLSLIACLLASTQAVALSTDRQPV

```


Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

FEATURES

source

Location/Qualifiers
1..13836

/organism="Xylella fastidiosa 9a5c"
/db_xref="taxon:160492"

/clone="9a5c"

/complement(195..803)

/gene="XF2101"

/complement(195..803)

/gene="XF2101"

/note="similar to G11652989 (percent identity: 30 %/query alignment coverage: 87.6 %/subject alignment coverage: 67.6 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF84900.1"

/db_xref="GI:9107229"

/translation="MSRCITGRMOWICVYVVLVFMVLVAVAGNIWIGORDOASS
DVIVGAAYDAKPSVPEERIRHGLYLQHGVAPTLFTGGSSARFESQV
RALRDYPPRALIETVSRTROMDQARQLMRYGLRRALVSDPLHMAARLCR
EIDIDALASATPSTRSLHRTMRFLLEVFYHDDVIOAN"

/complement(800..1042)

/gene="XF2102"

/complement(800..1042)

/gene="XF2102"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF84901.1"

/db_xref="GI:9107230"

/translation="MARVIMFYSFALSVKSCIGVTCMLLHAYVFQGLGFEFRAL
CCDVAVLCHVDISARLSCGCSAGFFICVSLVVS"

/complement(1397..1732)

/gene="XF2103"

/complement(1397..1732)

/gene="XF2103"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF84902.1"

/db_xref="GI:9107231"

/translation="MKARLFLPILMSATSCGATYVSNQYVHGPTVEVYHPPNYY
PHGGMHDLFNPPIPRDFYRTHLHLYWQMRPHDPSRPLPKQPIKQOPNHIE
MGIRDKQPI"

/complement(1914..2024)

/gene="XF2104"

/complement(1914..2024)

/gene="XF2104"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF84903.1"

/db_xref="GI:9107232"

/translation="MNITADODMLLCSLELMKICPHCVDTLKHQRLAR"

/complement(2067..2225)

/gene="XF2105"

/complement(2067..2225)

/gene="XF2105"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF84904.1"

/db_xref="GI:9107237"

/translation="MAAF84904.1"

/complement(1397..1732)

/gene="XF2106"

/complement(1397..1732)

/gene="XF2106"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

gene

CDS

/db_xref="GI:9107233"
/translation="MAASREEMAFYVPCVIFRNALFVGFYAVDVAPPAALRL
QAOSSGLE"

2271..4160

/gene="XF2106"

2271..4160

/gene="XF2106"

/note="similar to SPI25756 (percent identity: 65 %/query alignment coverage: 98.6 %/subject alignment coverage: 98.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="glucose inhibited division protein A"

/protein_id="AAF84905.1"

/db_xref="GI:9107234"

/translation="MTDSFYGYDVIYIGAGSSEALAAARGANTLLTINIEITG
AMSCNPAITGIGKGLVKEIDALGMAHAADAAGIOWRLNLSKGAVARATRCQDR
ALYRAAIROMIEGQARLNIQAEVDDLFEGDYVCGAITHTGHLFAPAVVLTAGFL
AGRIHIGPTVOAAGRMGDPATMLAARLEFLVARLKGTPPRIDGRTLYTAMQE
OPGAPRPTKSFIGNSASHPROVSCWITQRTERTHEIIRALRSPYSGQIEGAPR
YCPSEEDVYVAFKNSHQIFVEPEGINVIDIIPNGISTLPDVOLELRSTRGEDQ
AHITRPGVALEYDEFDRGLKASLETIAIGLEFAGQINCTGTGEEAAAGLIGLNA
ARHVGCLSWTPRRDQAVLGVVDLITHTGNEPRTMFTSRAEYRLQREDNADRLT
AICGDLGILIDARAHFNAKOEAARCEGSLAEMAPNAGALREVEYETLGLSRET
NITDLMKRPEDVYALMRVPSLGPVDQAOVAQVEISVAYGLNRQSEITROORH
EATNPLEFEEYAAVARGLSTEVLOKQIHQIPQYVQAGRIQGMPTAISLLVHLNRR
RNRYA"

/complement(4348..4560)

/gene="XF2107"

/complement(4348..4560)

/gene="XF2107"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF84906.1"

/db_xref="GI:9107235"

/translation="MFILMCVHTHPLIDELVEAMRAGPVIRSMYNAIKODGSSGITE
RLPVRFWSGRASKIDATAAAILASV"

/complement(4928..5001)

/gene="XF2108"

/complement(4928..5001)

/gene="XF2108"

/note="found by tRNAscan"

/product="tRNA-Cys"

4928..5001

/gene="tRNA-Cys"

5207..5392

/gene="XF2108"

5207..5392

/gene="XF2108"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF84907.1"

/db_xref="GI:9107236"

/translation="MQADGTTIGLIDISTKRWATAPLSGONPLSIDKSDPVLGALR
NSAKNDTQRLAPLITA"

/complement(5566..5892)

/gene="XF2109"

/complement(5566..5892)

/gene="XF2109"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF84908.1"

/db_xref="GI:9107237"

/translation="MAAF84908.1"

/complement(1397..1732)

/gene="XF2110"

/complement(1397..1732)

/gene="XF2110"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF84909.1"

/db_xref="GI:9107237"

Query Match

64.0%; Score 16; DB 1; Length 13856;

Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 gatcgctcttgaag 23
|||||
Db 2652 GATCGCTGTTGAAG 2667

RESULT 42
AE004580 15247 bp DNA BCT 30-AUG-2000
LOCUS Pseudomonas aeruginosa PA01, section 141 of 529 of the complete
DEFINITION genome.
ACCESSION AE004580 AE004091
VERSION AE004580.1 GI:9947468
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 15247)
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
Hickey, M.J., Brinkman, F.S.L., Huftagle, W.O., Kowalik, D.J., Lagrou, M.,
Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y.,
Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Lardig, K., Lim, R.,
Smith, K., Spencer, D., Wong, G.K., Wu, Z., Paulsen, I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
2 (bases 1 to 15247)
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
Hickey, M.J., Brinkman, F.S.L., Huftagle, W.O., Kowalik, D.J.,
Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E.,
Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
Folger, K.R., Kas, A., Lardig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sifer, M.H.,
Hancock, R.E.W., Lofly, S. and Olson, M.V.
Direct Submission
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
TITLE
JOURNAL
FEATURES
source
1. 15247
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
complement(122..640)
/gene="hc"
/note="PA1512"
complement(122..640)
/gene="hc"
/codon_start=1
/transl_table=11
/product="secreted protein Hcp"
/protein_id="AAG04901.1"
/db_xref="GI:9947469"
/translation="MATPRYMSITGKGLITAGAFTEDSGNTVQESBEDQVWQGS;
NEVILIPDPSGQPTGPQVHKPVVITKFEKASPLLAALISGERLTKVELQWRTS
ASGDQEHYTTVLEDAITVIDIKDYMHNCODPGNAHPTHELDVHFYRKITMTHEVSGT
SGSDHRSVPVAC"
complement(1126..2403)
/gene="PA1513"
complement(1126..2403)
/gene="PA1513"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04902.1"
/db_xref="GI:9947470"
/translation="NEAHLIEVLNLIVKIMIVGTAMIGASFYFVLENNLNSNPK
FGLSGDLMAIHGGIYHLEKYLAPKMEENLHMKRYAVSTWMSGVLLITVLYLN
TLTILPQSDLAFAAAIAIGISLVCGWFVYITLQDSALGKTPALLGLVFLVAA

gene
complement(2452..2961)
/gene="PA1514"
complement(2452..2961)
/gene="PA1514"
complement(2452..2961)
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG04903.1"
/db_xref="GI:9947471"
/translation="MRTIKIEPLTKEAFAPGDIETAGSDYPMINNGSTRYHKLAT
VETAPEDNATIIISFSAKLEMLPLRIMRLEPILGSOAPILIGNPETLYVAVLGPV
VPGIVRATITNGRCGVNTHRCVWHHPVLITTEKRDPLVYBRSGSGNCDHEFTEDBO
LLDPDSNO"
complement(3016..4014)
/gene="a1c"
/note="PA1515"
complement(3016..4014)
/gene="a1c"
/codon_start=1
/transl_table=11
/product="allantoicase"
/protein_id="AAG04904.1"
/db_xref="GI:9947472"
/translation="MNAHSPFRHYDLADARLIGSOVAVASDEMPAPSMIQAQEPV
WKEGVDPDSCGKMDGEMETRRKREGHQAATRLGVPYLVKGVYIDIRFTFGNPPAS
LDGFCACAGDDSDTSVSEVLAAGLGDSDSHHTPIIDDERPMHLRNITPDGTAFL
RLVGPYRDMSNOPPGTALDIAAIVNGRSLACSDHFGMGLNLNGRAIINGDME
TGRRRTPGHDPVVALGHPGSIKAAVVDITLHFGNVPESCSIGAAFEVGENARIBAO
SLFWRELIPQKLEHMOHREHFRHNLALCPITVRLNIIPDGVSRLLEGRQDL"
complement(4032..4547)
/gene="PA1516"
complement(4032..4547)
/gene="PA1516"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04905.1"
/db_xref="GI:9947473"
/translation="MSRFQTLTPASLSREAFIEADYHSEWPAEKAYDGLDLSL
NDLEGLQRMADILSASREAOALINAHPLDLAGKAIRBELTASSTSEAGAGIHGC
SAEEFARFTEELNDAYKARFGFPIKAVKGSNRHQTIAAFERIQHSADDEFATALEI
NKILPRLQOL"
complement(4544..5470)
/gene="PA1517"
complement(4544..5470)
/gene="PA1517"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG04906.1"
/db_xref="GI:9947474"
/translation="MSADYPRDLIGYGNPPIHWPBGDARIALSVLYNEEGERCVL
HGDKESEAFLESEVAAQPILOGVHRMSMESLYEYSGRAGWRLLKTKRRNVPLTFAV
AMAAQRNEVIRAMVADGHEICSHGYRIDYQYMDAOREHMLLEAIRITELTGORP
VGWVTGRTGCTNTRRLVWEGGLYDSDTYDQDLDLDPYNDPASTAKRPLVLPYLDNDM
RFOVQSGFNNGBOFOYLKAPRVLYEPGATTAKMLSLGICHLRLIGRAPMAALERFI
OYASDHKVMFARREDIARHREHFPQTEA"
5843..6223
/gene="PA1518"
5843..6223
/gene="PA1518"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG04907.1"
/db_xref="GI:9947475"
/translation="MSASGPOPGRGLTTHVLDSAGHCPGHGKITELYVEGOOLEI
ATTLVNHNDGCDQPVIGDEDFRCGVYQVLFMAGDYRANGVOLPRAPFLDQVYVLRGI

```

gene      ASADHVAHPLLISPYSTYKGS"
CDS       complement(6464..7813)
          /gene="PA1519"
          /complement(6464..7813)
          /gene="PA1519"
          /codon_start=1
          /transl_table=11
          /product="probable transporter"
          /protein_id="AAG04908.1"
          /db_xref="GI:9947476"
          /translation="MESTKOEQGIYAATPPATGLIERFLKQHGTTVTELAGLT
          PFTMAVITFVNINIMADGIDHGAAPVATCAALAGCELMGLVANNMVGLAGSLN
          AFFTYVTCSTKYSNOAIGAFTSCWPMFLTPSRREMILNIPRLFRPMGAGC
          LFGLGLGTAGTIVASPATIKLITKCHLITSPGLALCFMLAVETKRVKGLIST
          LSVTLVGMALGLVOYGVFSAPPPLAPFLANDIAGAFNVTMI SVTLAEVHMDTA
          GTLMGVAORHLVKEDEGRLENLSKAMKADSASVGMIGVPTVSVESAAGAAG
          RGLTAVVGVLEFVAAMPAPLAGMIPAFATGALIVAMLMGMAHIDDEHETI
          PAIVTYMPLTFESVADGIALGFTIYVAMKVLTKHREVSASLVAFVIFVAKFTFL"
          8188..8967
          /gene="PA1520"
          8188..8967
          /gene="PA1520"
          /codon_start=1
          /transl_table=11
          /product="probable transcriptional regulator"
          /protein_id="AAG04909.1"
          /db_xref="GI:9947477"
          /translation="MTDQLQHRKRPRTGKARSGTODEIYVAHIFDAILEGRAPGT
          KSEEMIGEITGVSRITIRRALSRUAEQVLLRPNRGAVVASPSIDEARQILLFART
          VERATLELATDNATAERLAELEMYVQESFARGDRCAGITLGEFHLKIAEMRNA
          PLYSPORSIVSOTSLITAOYESGSRGSCFDEHNEITLDAIEGDKERAVALTMIMHMEH
          IOSKINLESNDNSGDLHAFSLGCKKKPRAKKADSDSA"
          complement(9021..10325)
          /gene="PA1521"
          complement(9021..10325)
          /gene="PA1521"
          /codon_start=1
          /transl_table=11
          /product="probable guanine deaminase"
          /protein_id="AAG04910.1"
          /db_xref="GI:9947478"
          /translation="MTRNAKAYRAALIMSIADPAEYGVRSYEYEDGLLVEDGKVA
          RLDGATTLGGEIEVEVEFYEDALITPGFIDTHIHPQGMISASGEOLIMVNTYTE
          PPRRGDOAHADVOAEITFLOELIRNGTTALVPGVHQSVESELEAARRDLRLIA
          GKYMARNAPDYLITPBAESSYRDSKALIRRMGCGLLAAVPRFAPSTABOLMAA
          RLUREHPGYLTHHSENLKEIWEKLELPERSGTIDYVDHGLGPRSVFAGCHYC
          DEGCORLAETGSVAVAFCPISNLELGSGLFDLPKLEKRYKRVGIGTVGAGTSFSLOS
          LNEAYVMOLOGARLDPFKSLYATLGARALELDRIGSFATSNDEADVLDYHATP
          LLSYRLSOGASIAERLEFALTITLGDRTVKEFPAAGRSVHRD"
          complement(10395..11234)
          /gene="PA1522"
          complement(10395..11234)
          /gene="PA1522"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAG04911.1"
          /db_xref="GI:9947479"
          /translation="MMWISALAELOSGEPVUVTIIEERGSPRNAGSKMVAARL
          YDTIGGHELEYSRAQALAREMLAARTQDRLERSLSIASLCOCCGATVLLFERMGQPO
          AHIAVFCAGHAGALVPLASLPCRVWIDISREHERPALLPDGEKVVNDVEVLDVEER
          MPGSHPTIYVTHHPLDLELTALILARNDFAYGLGSKTKRVKFEHRLREKGVDER
          LOBRMCPMGLEQYKXKLPALIAISINGEIVATYNAFGLERKKGSSVTRTPESRRA
          HES"
Query Match      64.0%; Score 16; DB 1; Length 15247;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 43
HS3164
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
Source
The true right end of clone 314G4 is at 31557.
314G4 is from a 280kb clone contig extending from the telomere of
16p. Higgs D.R., Flint J. unpublished. MRC Molecular Haematology
Unit, Institute of Molecular Medicine, Oxford.
314G4 came from the Los Alamos, flow sorted human Chromosome 16
libraries constructed by Norman Doggett (unpublished).
location/Qualifiers
1..31557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
/clone="LA16-314G4"
/clone.lib="LA16"
428..716
/feature="Alu repeat: matches 308..1 of consensus"
914..1691
/feature="expressed region - matched by multiple ESTs"
1692..2316
/partial
/feature="match: 5' EST R51457 clone 39198"
2290..2810
/partial
/feature="match: 5' EST T87094 clone 115315"
complement(3171..3720)
/feature="expressed region - matched by multiple ESTs"
5136..5489
/feature="3 copies of 118 mer 88 % conserved"
5908..6086
/feature="match: EST Z39463 clone c-1ag06"
7536..7567
/partial
/feature="Alu repeat: matches 1..32 of consensus"
7568..7856
/feature="Alu repeat: matches 1..308 of consensus"
9446..9734
/feature="Alu repeat: matches 308..1 of consensus"
12398..13777
/feature="putative Cpg island"
13241..13275
/feature="7 copies of 5 mer 86 % conserved"
14561..14677
/partial
/feature="Alu repeat: matches 254..136 of consensus"
14680..14971
/feature="Alu repeat: matches 308..1 of consensus"
14974..15106
/partial

```

```

repeat_region      /note="Alu repeat: matches 145. .1 of consensus"
15127. .15231
/partial
/note="Alu repeat: matches 116. .1 of consensus"
15248. .15535
repeat_region      /note="Alu repeat: matches 308. .1 of consensus"
15607. .15917
repeat_region      /note="Alu repeat: matches 308. .1 of consensus"
16005. .16114
/partial
/note="Alu repeat: matches 108. .1 of consensus"
16123. .16283
repeat_region      /note="Alu repeat: matches 308. .150 of consensus"
17661. .18594
/note="putative CpG island"
18912. .19019
repeat_region      /note="3 copies of 36 mer 89 % conserved"
19514. .19621
/note="MER22 element fragment"
19922. .20499
misc_feature        /note="match: 5' EST H21176 clone 173392"
20504. .20539
repeat_region      /note="4 copies of 9 mer 97 % conserved"
21472. .21766
repeat_region      /note="Alu repeat: matches 308. .1 of consensus"
23040. .23378
misc_feature        /note="expressed region - matched by multiple ESTs"
complement(25620. .25910)
misc_feature        /note="expressed region - matched by multiple ESTs"
26707. .26817
/partial
/note="Alu repeat: matches 298. .184 of consensus"
26818. .26963
repeat_region      /note="Alu repeat: matches 160. .1 of consensus"
complement(27039. .27209)
misc_feature        /note="expressed region - matched by multiple ESTs"
30474. .30611
/note="3 copies of 46 mer 94 % conserved"
31087. .31220
misc_feature        /note="expressed region - matched by multiple ESTs"
31496. .31557
/partial
/note="Alu repeat: matches 1. .62 of consensus"
BASE COUNT      5713 a 9969 c 9918 g 5957 t
ORIGIN
Query Match      64.0%; Score 16; DB 9; Length 31557;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 gttgtgctgctgctg 16
|||||
Db      10104 GTTGCTGATCTGCTG 10119
|||||
RESULT  44
SC8520X      41200 bp      DNA      PLN      11-AUG-1997
LOCUS      S.cerevisiae chromosome XIII cosmid 8520.
DEFINITION      249705.249700.271257
ACCESSION      249705.1 GI:825556
VERSION      249705.1
KEYWORDS      ALD2: aldehyde dehydrogenase; ATPase; CBF3b: centromere DNA-binding
protein; coiled coil; DNA mismatch repair; DnaJ: Mitochondrial
carrier protein; MLH1; Smp2: transfer RNA-Met; transfer RNA-Val.
baker's yeast.
SOURCE      Saccharomyces cerevisiae
ORGANISM      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE      1 (bases 1 to 41200)
AUTHORS      Hunt, S. and Bowman, S.

```

```

JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 41200)
AUTHORS      Barrell, B. and Rajandream, M.A.
TITLE      Direct Submission
JOURNAL      Submitted (19-MAY-1995) Saccharomyces cerevisiae chromosome XIII
Sequencing Project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1RQ E-mail: barrell@sanger.ac.uk
Notes:
All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50%
of their length by a larger CDS have been omitted from this
analysis.
Details of the omitted CDS are available on request. The more
significant matches with motifs in the PROSITE database are
also included but some of these may be fortuitous. The length in
codons and the calculated codon adaptation index (CAI)
is given for each CDS.
Cosmid 8520 is overlapped at the start by cosmid 9375, EMBL ID
SC9375,
accession no. 247071 and at the end by cosmid 8010.
Location/Qualifiers
1. .41200
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 8520"
/map="13R"
1. 365
/note="overlap with cosmid 9375"
<1. .514
/note="Ym8520.01, len: > 170, CAI: 0.25, overlaps an
extends Ym9375.22, SW:YMAL.YEAST P28625, previously noted
hypothetical 41.6 KD protein in IMP1 3' region (RF1095),
hypothetical transmembrane protein"
/codon_start=2
/product="unknown"
/protein_id="CAA89788.1"
/db_xref="GI:825557"
/translat="SWISS-PROT:P28625"
/translation="VVGISKNSIERKRLGYDLYPVPDESGIVENVKKAKOSYLEND
KEDMFDVSNHDFPVIOFLPKRKNSEFYVPTGINNKAADPKNSMRPFVLSITLK
A1NPFKKYNNRFRHPYPPNNFIEVGNEMIKGTYKPPIDSVYEFQYKADIRLMSNR
AKGVVYKKK"
756. .2183
/note="Ym8520.02, unknown, len: 475, CAI: 0.12"
/codon_start=1
/product="unknown"
/protein_id="CAA89789.1"
/db_xref="GI:825558"
/translat="SWISS-PROT:Q03790"
/translation="MADIQKENSRRFTNVSVIAPESQGHQKQEQLEQKQPTG
LRGLNGPPSADQPLEMEDPSTVSGELDNKPNPNNPKRAIPNSIIRKSGSLSP
VRSDSADVPAPFNSNGNVTFGSKKDPRLIKVSPDNNSNNNAHSSDICTGVFDS
NEAPRTSLADQKEDGTFSSKRDNTEDPNSNTTPCKPPTATSPRPLEKTSRL
NEFDKNTKTPNTASSEASAGSKEGASTWMDRAIITIGYPTIANSITILHANGEI
LEDFRVIKDFKLNKSNKSKSPSLTAQKPYITYGQWKLTYKSELKSRALQENGII
MNGTLLGCVSYSPALKQLASLKSEETINNKTSSOTSSKSLNSYKRTBIEPKAK
AKAVTSKVNAAEFKSKNSTSFKNPRLIEIKGRSLFLNRCKIHSGVLSIESDLK
REOASKSKSKSWLNRLNNMLFGMNDL"
complement(2320. .4503)
/note="Ym8520.03c, unknown, len: 727, CAI: 0.12, some
similarity to WP:10448.16 CE01070 calpain thiol protease,
(27.1% identity in 155 aa overlap)"
/codon_start=1
/product="unknown"
/protein_id="CAA89790.1"
/db_xref="GI:825559"
/translat="SWISS-PROT:Q03792"
/translation="MNDHFRNAATKSIYCNAEDSSIIINRLVGLANKSEDSTFIEA
VLVLKENSRYKQKLFMTSTINSRPPPISEASPVSNKNTKCYCAPGTELDQR
YGRAKLQNEEDYSGIEGCRDVPDCSLVASLIINRSKLNLPILIKQISSTKYHNS
FNSNKRLLVTVDISQIPTSVDGKQLSKNSIDSKIGELALLVSKGYSTDGSNISI

```

DTVRLSGFLPEITGVNSYPFEKLMKFKSNLCIMAGTGNBSNDMIKPLVENHDYII
 DITVRLVYKLRDPRNSALNVEISYEDYLMKFKOLYINMNOEKLFRKRSQVHFPEYDS
 RYKRSYADKPLPLVNLKSKVTERVWILLESHELODBSGDNRKSVSEINREPECITCP
 IEPVECGGNHIGLOLVKLRDLDAETERLYCYSTNNFSTHSFVKEICFQKLDIT
 KSLFAKVFSEPPYEIEGASFDQCNFQNPTELEHSEDOYQVLMADAACSTSHDL
 INIOVYVNDYELKIPIMEDNHOPOGOLKODVILTNVKYIACSTGIPASTFEL
 LASILRSSMRLISGILTSVNLVLIYGTGYHGRHMRHFKETSDKLIDMTLPTKYST
 NKLFRVYVSSARLRMRKCNIFEPELALCYHCEOEKPTCSGSDKLIDMTLPTKYST
 VLMIRSVPISSCLPTEGDLDELFFVSSOKIRIKEXSDVITPK"
 4872..6515
 /note="Ym8520..04, unknown, len: 547, CAI: 0.10, potential
 integral membrane protein"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA89791.1"
 /db_xref="GI:825560"
 /db_xref="SWISS-PROT:Q03795"
 /translation="MKRHNQSKMGNTNHPGHLKSPGVGNVVALAGTPVLESEYAPQ
 LLSKCHIVSASKLSLSLTIGSLMGLAGITVDRSPKLSCLIGSKCVFAYILNL
 CYKHEMSFPLISLVLIGYSGVSGEYASVCANTNFPQHRGTAGAPVLYGLSGM
 VESYLSKLGEGNIEHVFIFLIMAGCCMLIYGPSLIDFSAEGDDASIKRMELOKSR
 ETDNIVLPLENSNDYICSPVRSPPATYETATSDNFQETSEFFALEDRRLRPL
 SSPSPHTVYDEDENTSKNTVGENSAQSMRLHVFQSLKSTFIGYIVLIGLVGL
 MYIVSGMVQAVQSTPLNOLPINAETQISLOVTLISLFCGRSSGPISEFLVK
 FKAOLMNIIVIASLVLASNKISHDESSIEDPSLAKSKFNISVCAITGYSEGLV
 FGTSPVADRGNGSYTLMGVLTGTGVSFVSPTDILGRDFKANTGDDGCKKGV
 LCVSTFWVTKYCAFNILFYILIGITTYRRRATANSI"
 complement(6621..7337)
 /note="Ym8520..05c, unknown, len: 238, CAI: 0.15"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA89792.1"
 /db_xref="GI:825561"
 /db_xref="SWISS-PROT:Q03796"
 /translation="MSHKULILPLIKETPKFPOSIDHDEGLNVAADLDTIILKK
 SPNISFSRASDMQFINENSKSTLDYLCNIDNPVAVIIVFSGQGVITVPRIKRS
 CKRYNKLILFLKAIKDEGETLSHRLVLAAPRPRTFAMNSKIFPAISGTSYN
 DNPIFEKVKPMTGWEFFERDLSEAVYSEQISPKIMLIYCGDAAGRRKKDSDSD
 IKFAENLHYEKPPEIFHG"
 complement(7598..8365)
 /note="Ym8520..06c, unknown, len: 255, CAI: 0.17"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA89793.1"
 /db_xref="GI:825562"
 /db_xref="SWISS-PROT:Q03798"
 /translation="MLRLKSVLASCHNCKEVCGLDPOKOLPUEFPLLRARYSTO
 SSTKRSKSDKIDAGFKRIFLVAITGVIVFQVQSLDKKPKTTLSSEEFENVKQ
 LKRRVAIFQGEVDIKFSLSPSIEETRYVLOSGDDINELOFVDPVAVIYRTLRD
 DRYEALNLYKKGCDYANLPTGMVLMGLGRYFKNPFTGDKLVAVNPHSLADA
 TPEENEVSIVSKIFPRKLSGSDVCKYETVYKADIT"
 8569..9036
 /note="Ym8520..07, unknown, len: 155, CAI: 0.13"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA89794.1"
 /db_xref="GI:825563"
 /db_xref="SWISS-PROT:Q03799"
 /translation="MSLVKLANCAHLNCKSVKVALTSIPYTKLOLQFAYNLVQOGF
 LSSLOGSTMGPDKDFEVETPDNISTRIAMWLKTRDKRPVLSCKLISKNRSRHLRP
 MEDMKKICSGVIRINIKPLOGEILLVHNNHMDINAIKSLDGEVLCKRK"
 complement(9204..9276)
 /note="crna-met"
 /gene="crna-met"
 /note="crna-met gene, anticodon cat, len: 73"
 complement(9204..9276)
 /gene="crna-met"
 complement(9432..9659)
 /note="remnant of solo delta element"
 complement(10797..11249)
 /note="Ym8520..08c, unknown, len: 150, CAI: 0.13, possible
 coiled-coil protein"
 /codon_start=1
 /product="unknown"

/protein_id="CAA89795.1"
 /db_xref="GI:825564"
 /db_xref="SWISS-PROT:Q03818"
 /translation="MGNFIIITERRKKKERSNPQDTSMDLLIRRLTRDNKEAHUNE
 LFQDNGATIGNIVSHDDLNLNLALQELKSEOEIRRLKEVIALKKNTERLUNDE
 LISGIIENNVLQOKLSDLKKEHSQLVARMKTKETKEAMNSRIDGK"
 11387..13837
 /note="Ym8520..09, unknown, len: 816, CAI: 0.13"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA89796.1"
 /db_xref="GI:825565"
 /db_xref="SWISS-PROT:Q03823"
 /translation="MDVVOOKODKOLHOGFOEQOJIREDOQEVPPQRPQNRKPKPM
 NSTADDEPNRTIAEYVNGGRRSPPTDODSDNSNDKGIKSIASFATSRYSRAP
 IYVDNTRYSQLNTBOINLENAKOLISKSKSKWYAIPIHSSNMIDSLDIPG
 IISVSGTSKACPLPLNKPFGEGNGVNFINDSLILPSDPLNLAHQPLATKVLN
 TIKNYNFPNEOHLYLQKKATLAKORIIISVVDLPEKYEORLESQSAAYLSR
 KLSQNLAOEQPOVULILSFQCLINODLIPYKECVELLNMAHLKEVDYSIFGVYV
 HSYPLTLAKYIYONNEVEFEDENTVSVLSFOSLOGYREMDHSTFSSNYNLG
 SNSNTENDSDHDSNDNFTTKSOKOLFOGIDKKOQDTLSKIKNYRIDSSSK
 LVODALDMLFNMDDPRLTRFEGKLYONPRTISKLAIDVNHPRILANKNGKMGID
 LKANNNINLITDEATSNIDVAVRTPNEPSRLKIFPNRLEITLMDILITENLGK
 QETPLILSPFISRFENDYTLPPNIRKQYQNSKIMQEMDSKMKMGHOLDNDR
 EGESLGSSESLPENISTYKDFLOEVYQONKESDFVRSYIDYDDKVLPOEILNT
 IFMKNPLSRKHLRINDIDLPSTILANTVNOYDLMVKIHDSFSLQKMLPOEIPHA
 LRLSINCRLDSTSGSPVQPOQDIYEALRLRLTEIMRYQDMSPTPLGLHLRLIS
 VLAINDPKNLINDVRRT"
 14039..14713
 /note="Ym8520..10, putative DNAJ gene, len: 224, CAI: 0.14,
 similar to DNAJ proteins"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA89797.1"
 /db_xref="GI:825566"
 /db_xref="SWISS-PROT:P48353"
 /translation="MSFTEDQKIALEILSKDHEFEYELKVDKATDSEIKKAYRKL
 AIKHPDKNSHPKAGEAFKVINRAFEVLSSEKRSYDRIGRPDPRDORGAASGF
 RGSAGSGPMGSGFEDMFNRSFGORGPREDIFELFVNGSGSPFSGSPSPASPS
 FCGPGCFRVYTNNGSGSPFRQQRPSRQOQQAENNAVSQLKNMLVLIPIIVLPMI
 KDYLFS"

Query Match 64.0%; Score 16; DB 8; Length 41200;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgacatgcgtcgt 18
 Db 25021 TGGTGCATCTGCTGTT 25036

RESULT 45
 AX008043/c 46340 bp DNA PAT 06-SEP-2000
 LOCUS AX008043 Sequence 66 from Patent WO967395.
 DEFINITION AX008043
 ACCESSION AX008043
 VERSION AX008043.1 GI:9995728
 KEYWORDS
 SOURCE
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 46340)
 Perin,J.P., Rieger,F. and Alliel,P.M.
 Nucleic sequence and deduced protein sequence family with human
 endogenous retroviral motifs, and their uses
 Patent: WO 96/7395-A 66 29-DEC-1999;
 INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
 FRANCOIS (FR); ALLIEL PATRICK M (FR)
 FEATURES
 source Location/Qualifiers
 1..46340
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 16104 a 8737 c 8435 g 13064 t
ORIGIN

Query Match

64.0%; Score 16; DB 6; Length 46340;
Best Local Similarity 100.0%; Pred. No. 92;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gctgacatcgtctt 19
|||||
Db 1254 GCTGATCTGCTGTT 1239

Search completed: December 26, 2001, 12:50:57
Job time: 8779 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:16:32 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196F-3

Perfect score: 25
Sequence: 1 gttgtggtatctgtgtgttgagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 58608

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

1: N.Geneseq_1101.*
2: /SID22/gcgdata/geneseq/geneseqn/NA1980.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn/NA1981.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn/NA1982.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn/NA1983.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn/NA1984.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn/NA1985.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn/NA1986.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn/NA1987.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn/NA1988.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn/NA1989.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn/NA1990.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn/NA1991.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn/NA1992.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn/NA1993.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn/NA1994.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn/NA1995.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn/NA1996.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn/NA1997.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn/NA1998.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn/NA1999.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn/NA2000.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA01303 E. coli biotin syn
3	25	100.0	1084	10	AA01329 E. coli B gene
4	25	100.0	1121	7	AA060496 Sequence encoding
5	25	100.0	5872	15	AA062386 Biotin-biosynthesi
6	16	64.0	5000	21	AA236139 Nucleotide sequenc
7	15	60.0	1683	16	AA014850 Maize U14.4 DNA se
8	15	60.0	1683	16	AA014850 Maize U14.4 DNA se
9	15	60.0	1689	20	AA223024 Rat kd312 polypept
10	15	60.0	2022	21	AA041132 Zea mays DNA fragm
11	15	60.0	2034	20	AA04506 Maize genomic c1on

C	12	15	60.0	3079	20	AA223025	Rat kd312 genomic
C	13	15	60.0	6703	19	AA049536	Adenylate cyclase
C	14	15	56.0	275	10	AA212718	Human gene express
C	15	14	56.0	336	22	AA05487	Novel human polynu
C	16	14	56.0	379	20	AA07319	EST clone BP797
C	17	14	56.0	383	21	AA026528	Human secreted pro
C	18	14	56.0	384	18	AA010888	Human ORK27 gene
C	19	14	56.0	393	22	AA010789	Human cDNA clone (
C	20	14	56.0	464	22	AA10569	Probe #502 for gen
C	21	14	56.0	464	22	AA131822	Probe #492 used to
C	22	14	56.0	464	22	AA100501	Probe #2833 used t
C	23	14	56.0	469	22	AA112900	Probe #2944 used t
C	24	14	56.0	469	22	AA134258	Probe #2810 used t
C	25	14	56.0	469	22	AA102819	Probe #12072 for g
C	26	14	56.0	476	22	AA122139	Probe #16121 used
C	27	14	56.0	476	22	AA147435	Probe #7830 used t
C	28	14	56.0	476	22	AA107839	Probe #9754 for ge
C	29	14	56.0	558	22	AA119821	Probe #13702 used
C	30	14	56.0	558	22	AA145016	Probe #5528 used t
C	31	14	56.0	558	22	AA105537	Probe #8797 for ge
C	32	14	56.0	575	22	AA118864	Probe #12672 used
C	33	14	56.0	575	22	AA143986	Human secreted pro
C	34	14	56.0	612	21	AA099862	Human secreted exp
C	35	14	56.0	619	21	AA044124	Aspergillus oryzae
C	36	14	56.0	707	21	AA014018	Mouse antibody F4-
C	37	14	56.0	735	16	AA092502	Nucleotide sequenc
C	38	14	56.0	753	20	AA0200347	Human FATP3 coding
C	39	14	56.0	753	22	AA089005	Trichoderma reesei
C	40	14	56.0	769	21	AA015012	Human secreted pro
C	41	14	56.0	891	20	AA0206233	L-asparaginase DNA
C	42	14	56.0	1044	21	AA082834	E. coli S-adenosyl
C	43	14	56.0	1155	20	AA0209784	Met K oligonucleot
C	44	14	56.0	1155	22	AA099433	Nucleotide sequenc
C	45	14	56.0	1173	22	AA0200358	Human FATP3 partia
C	46	14	56.0	1173	22	AA089020	Arabidopsis thalia
C	47	14	56.0	1220	21	AA044861	N. meningitidis pa
C	48	14	56.0	1367	21	AA081673	Human ORK27 gene
C	49	14	56.0	1493	18	AA001089	Flax 2S storage pr
C	50	14	56.0	1676	22	AA003029	Human polynucleoti
C	51	14	56.0	1699	22	AA158178	Chimeric gene cont
C	52	14	56.0	1848	18	AA096346	Nucleotide sequenc
C	53	14	56.0	1998	20	AA0200367	Human FATP3 coding
C	54	14	56.0	1998	22	AA089041	Human cDNA sequenc
C	55	14	56.0	2080	20	AA030162	Human secreted pro
C	56	14	56.0	2087	20	AA0200343	Nucleotide sequenc
C	57	14	56.0	2087	22	AA089001	Murine FATP3 codin
C	58	14	56.0	2166	20	AA0200379	Nucleotide sequenc
C	59	14	56.0	2166	22	AA089041	Human FATP3 coding
C	60	14	56.0	2314	22	AA017621	Human cDNA sequenc
C	61	14	56.0	2319	22	AA019223	Human secreted pro
C	62	14	56.0	2339	22	AA007655	Human FATP3 coding
C	63	14	56.0	2402	22	AA089054	Human fatty acid t
C	64	14	56.0	2405	22	AA027417	Human secreted pro
C	65	14	56.0	2447	22	AA019186	Human PRO703 nucle
C	66	14	56.0	2574	21	AA0233977	Human PRO703 (UNQ3
C	67	14	56.0	2574	21	AA0233977	Human PRO703 nucle
C	68	14	56.0	2574	21	AA0233977	Human PRO703 nucle
C	69	14	56.0	2574	21	AA0233977	Human PRO703 nucle
C	70	14	56.0	2675	20	AA025020	Human H-beta 58 fa
C	71	14	56.0	2675	21	AA025020	Human H-beta 58 fa
C	72	14	56.0	2709	22	AA025020	Human colon cancer
C	73	14	56.0	2709	22	AA025020	Human colon cancer
C	74	14	56.0	2826	22	AA085785	cDNA encoding meta
C	75	14	56.0	2925	22	AA085785	cDNA encoding chim
C	76	14	56.0	3144	22	AA098356	Human EST-derived
C	77	14	56.0	3243	11	AA005686	Islets of Langerha
C	78	14	56.0	3243	20	AA023338	Human pancreatic i
C	79	14	56.0	3243	20	AA020919	Human islet cell a
C	80	14	56.0	3243	21	AA063557	Islet cell antilob
C	81	14	56.0	3282	16	AA046456	Pancreatic islet c
C	82	14	56.0	3421	22	AA080421	Human mcgUR5c cDNA
C	83	14	56.0	3591	20	AA014567	Human cDNA sequenc
C	84	14	56.0	3720	20	AA010801	Mouse metatrophic
C	84	14	56.0	3918	20	AA020970	Plasmid pHS1 metk
C	84	14	56.0	3918	20	AA020803	Rat metatrophic g

```

C 85 14 56.0 4085 16 AA080419 Human mGluR5a cDNA
C 86 14 56.0 4181 16 AA080420 Human mGluR5b cDNA
C 87 14 56.0 4207 22 AD040920 Human metabotropic
C 88 14 56.0 4207 22 AD05029 Human metabotropic
C 89 14 56.0 4207 22 AD05030 Human metabotropic
C 90 14 56.0 4207 22 AD05033 Human metabotropic
C 91 14 56.0 4303 22 AD04091 Human metabotropic
C 92 14 56.0 4303 22 AD05031 Human metabotropic
C 93 14 56.0 4303 22 AD05032 Human metabotropic
C 94 14 56.0 4566 19 AA041550 Nucleotide sequenc
C 95 14 56.0 4566 20 AA01089 Human G-protein co
C 96 14 56.0 4568 18 AA04039 Human G-protein re
C 97 14 56.0 4568 20 AA01094 Human G-protein co
C 98 14 56.0 4975 20 AA009792 Plasmid pHS1 mek/
C 99 14 56.0 5319 19 AA035235 R. prowazekii S-1a
C 100 14 56.0 7326 16 AA094456 Sequence encoding

```

ALIGNMENTS

RESULT 1

AA162941 standard; DNA; 839 BP.

XX AC AA162941:

XX DT 22-OCT-2001 (first entry)

XX DE Human genomic DNA SEQ ID NO 269.

XX KM Human: neurotropic; cytostatic; dermatological; virucide;
 immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnary;
 antiParkinsonian; antistickling; antianaemic; antiarthritic; cancer;
 antithematic; hepatotropic; cerebroprotective; antiinflammatory;
 antiallergic; antidiabetic; antileucic; anticonvulsant; antifungal;
 antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 neurologic disease; infection; nephrotropic; gene therapy; vaccine;
 ds.

XX OS Homo sapiens.

XX PN WO200155449-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001: 2001WO-US01346.

XX PR 31-JAN-2000: 2000US-0179065.
 PR 04-FEB-2000: 2000US-0180628.
 PR 19-MAY-2000: 2000US-0205515.
 PR 07-JUL-2000: 2000US-0216880.
 PR 14-JUL-2000: 2000US-0218290.
 PR 14-AUG-2000: 2000US-0223447.
 PR 01-SEP-2000: 2000US-0229343.
 PR 06-SEP-2000: 2000US-0230437.
 PR 08-SEP-2000: 2000US-0231243.
 PR 25-SEP-2000: 2000US-0234997.
 PR 29-SEP-2000: 2000US-0236367.
 PR 13-OCT-2000: 2000US-0239937.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 08-NOV-2000: 2000US-0246528.
 PR 17-NOV-2000: 2000US-0249210.
 PR 17-NOV-2000: 2000US-0249211.
 PR 17-NOV-2000: 2000US-0249214.
 PR 17-NOV-2000: 2000US-0249265.
 PR 01-DEC-2000: 2000US-0250160.
 PR 01-DEC-2000: 2000US-0250391.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.

```

PR 05-DEC-2000: 2000US-0256719.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251989.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.
XX (HUMA-) HUMAN GENOME SCL INC.
XX PA Rosen CA, Barash SC, Ruben SM;
XX PI WPI: 2001-476225/51.
XX DR
XX PT Novel plasma membrane associated proteins useful for diagnosing,
XX PT treating, preventing and/or prognosing disorders related to the
XX PT proteins, including cancer, immune response and neuronal disorders
XX PS Example 2: SEQ ID NO 269; 532pp + Sequence Listing; English.
XX CC The invention relates to novel genes (AA162752-AA162961) and proteins
XX CC (AA042347-AA042415) useful for preventing, treating or ameliorating
XX CC medical conditions e.g. by protein or gene therapy. The genes are
XX CC isolated from a range of human tissues disclosed in the specification.
XX CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX CC (d) wound healing; (e) neurological diseases e.g. cerebral ischaemia;
XX CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX CC and parasitic infections.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other:

```

Query Match 100.0%; Score 25; DB 22; Length 839;
 Best Local Similarity 100.0%; Pred. No. 0.00017;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 gttgctgactctgctgttgaagc 25
    |||||||
Db 89 gttgctgactctgctgttgaagc 113

```

RESULT 2
 ID AA01303 standard; DNA; 1041 BP.

XX AC AA01303:

XX DT 12-APR-1999 (first entry)

XX DE E. coli biotin synthetase (BioB) coding sequence.

XX KM DAP aminotransferase; diaminopelargonic acid; transgenic plant;
 biotin synthase; biotin production; vitamin H; BioB; ss.

XX OS Escherichia coli.

XX PN US5869719-A.

XX PD 09-FEB-1999.

XX PF 30-APR-1997: 97US-0846338.

XX PR 30-APR-1997: 97US-0846338.

XX PR 08-MAR-1995: 95US-0401068.

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Patton DA:
 XX
 DR WPI: 1999-152902/13.
 DR P-PSDB: AAW73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di:amino-pelargonic acid amino-transferase or biotin
 PT synthase
 XX
 PS Example 2: Column 37-40; 34pp; English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (Biot). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;
 SQ

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best local Similarity 100.0%; Pred. No. 0.00017;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtcgtgactcgtctgttgaagcg 25
 ||||||||||||||||||||
 Db 57 gtgtcgtgactcgtctgttgaagcg 81

RESULT 3
 AAN91329
 ID AAN91329 standard; DNA: 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KW E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 XX
 PM GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;
 SQ

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best local Similarity 100.0%; Pred. No. 0.00017;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtcgtgactcgtctgttgaagcg 25
 ||||||||||||||||||||
 Db 80 gtgtcgtgactcgtctgttgaagcg 104.

RESULT 4
 AAN60496
 ID AAN60496 standard; DNA: 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KW Biotin synthetic enzyme; E.coli; deshiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag=a
 XX
 PM JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB: AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp; Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing deshiobiotin.
 XX
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best local Similarity 100.0%; Pred. No. 0.00017;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtcgtgactcgtctgttgaagcg 25
 ||||||||||||||||||||
 Db 98 gtgtcgtgactcgtctgttgaagcg 122

RESULT 5
 AA062386
 ID AA062386 standard; DNA: 5872 BP.
 XX

PT for reducing gene silencing or inbreeding depression -

XX PS Claim 6; Page 36-38; 59pp; English.

CC The present sequence represents the DDM1 gene of *Arabidopsis thaliana*.

CC The gene is located on chromosome 5, lower arm, where it is flanked on

CC the centromeric side, within 20 kb, by a zinc-finger protein gene and

CC on the telomeric side, within 1 kb, by a glutamic acid tRNA gene. The

CC DDM1 protein is a member of the SWI2/SNF2 family of polypeptides. The

CC DDM1 protein is important for the maintenance of proper genomic cytosine

CC methylation, and its function appears to be necessary to maintain gene

CC silencing. The protein appears to function in the DNA methylation system

CC by affecting chromatin structure. Disruption of the gene is associated

CC with DNA hypomethylation. Four DDM1 gene alleles have been identified,

CC all of which are expected to destroy or severely reduce gene function.

CC The DDM1 nucleic acid can be used to create transgenic plants with

CC altered levels of DNA methylation. The mutant or engineered plants are

CC expected to be viable and display valuable characteristic such as

CC inbreeding depression or a reduction in gene silencing.

XX Sequence 5000 BP; 1386 A; 909 C; 1029 G; 1676 T; 0 other;

SO

Query Match 64.0%; Score 16; DB 21; Length 5000;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcctgacatcgcgtt 18

Db 1132 tgcctgacatcgcgtt 1147

|||||

RESULT 7

AA14850/C

ID AA14850 standard; DNA: 1683 BP.

XX

AC AA14850;

XX

DT 04-JUL-1996 (first entry)

XX

DE Maize U14.4 DNA sequence.

XX

XX snORNA: small nucleolar RNA; U14: non-translated sequence; conserved;

KW Box C; Box D; processing; pre-ribosomal RNA; ribosome formation;

KW normal growth phenotype; yeast; plant; maize; potato; 18S rRNA; primer;

KW PCR; polymerase chain reaction; amplification; ss.

XX

OS Zea mays.

XX

FT Key

FT misc_feature

FT /tag= a

FT /note= "nucleotides not identified in sequence"

FT repeat_unit

FT /tag= b

FT /note= "inverted repeat"

FT misc_RNA

FT /tag= c

FT /note= "M14.4"

FT misc_feature

FT /tag= d

FT /note= "Box C"

FT misc_feature

FT /tag= e

FT /note= "Box D"

FT repeat_unit

FT /tag= i

FT /note= "inverted repeat"

XX WO9530748-A2.

XX 16-NOV-1995.

PF 04-MAY-1995; 95WO-EP01694.

XX

PR 09-DEC-1994; 94EP-0119487.

PR 04-MAY-1994; 94WO-EP01409.

XX

PA (GENE-) GENE SHEARS PTY LTD.

XX

PI Brown JWS, Leader DJ, Waugh R;

XX WPI; 1995-404115/51.

DR

XX

PT Plant U14 nucleic acid sequences and derivatives - regulate

PT ribosomal RNA production and accumulation in plants

XX

PS Example; Fig 7A-D; 57pp; English.

XX

CC The present sequence shows the maize single U14 sequence U14.4,

CC including flanking intergenic sequences and inverted repeats, as obt.

CC by sequencing of fragments of isolated genomic clones. Inverted repeats

CC and Box C and D elements are shown above (see AAT08751-52). U14 genes

CC are small nucleolar RNAs present in the genome as non-translated

CC sequences. snORNAs are involved in processing of pre-ribosomal RNA

CC transcripts and ribosome formation. U14snORNA is required for a normal

CC growth phenotype and for processing of pre-rRNA transcripts in yeast.

CC Plant U14 sequences have been identified and characterised in the present

CC invention, and may be intron encoded.

XX

SO Sequence 1683 BP; 383 A; 387 C; 358 G; 548 T; 7 other;

Query Match 60.0%; Score 15; DB 16; Length 1683;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgcctgacatcgtt 15

Db 601 GTTGTGATGATCTGCT 587

|||||

RESULT 8

AAT08746/C

ID AAT08746 standard; DNA: 1683 BP.

XX

AC AAT08746;

XX

DT 04-JUL-1996 (first entry)

XX

DE Maize U14.4 DNA sequence.

XX

XX snORNA: small nucleolar RNA; U14: non-translated sequence; conserved;

KW Box C; Box D; processing; pre-ribosomal RNA; ribosome formation;

KW normal growth phenotype; yeast; plant; maize; potato; 18S rRNA; primer;

KW PCR; polymerase chain reaction; amplification; ss.

XX

OS Zea mays.

XX

FT Key

FT misc_feature

FT /tag= a

FT /note= "nucleotides not identified in sequence"

FT repeat_unit

FT /tag= b

FT /note= "inverted repeat"

FT misc_RNA

FT /tag= c

FT /note= "M14.4"

FT misc_feature

FT /tag= d

FT /note= "Box C"

FT misc_feature

FT /tag= e

FT /note= "Box D"

FT repeat_unit

FT /tag= i

FT /note= "inverted repeat"

```
FT      /*tag= f
FT      /note= "inverted repeat"
XX      MO9530747-A1.
XX      16-NOV-1995.
XX      04-MAY-1994; 94MO-EP01409.
XX      04-MAY-1994; 94MO-EP01409.
XX      (GENE-) GENE SHEPARS PTY LTD.
XX      Brown JWS, Leader DJ, Waugh R;
XX      WPI; 1995-404114/51.
XX      Nucleic acid contg. plant U14 sequences and their derivs. - for
XX      regulating prodn. of ribosomal RNA, also as stabilisers for
XX      heterologous RNA
XX      Example: Fig 7A-D; 57pp; English.
XX      The present sequence shows the maize single U14 sequence U14.4,
XX      including flanking intergenic sequences and inverted repeats, as obtd.
XX      by sequencing of fragments of isolated genomic clones. Inverted repeats
XX      and Box C and D elements are shown above (see AAT08729-30). U14 genes
XX      are small nucleolar RNAs present in the genome as non-translated
XX      sequences. SnRNAs are involved in processing of pre-ribosomal RNA
XX      transcripts and ribosome formation. U14snRNA is required for a normal
XX      growth phenotype and for processing of pre-rRNA transcripts in yeast.
XX      Plant U14 sequences have been identified and characterised in the present
XX      invention, and may be intron encoded.
XX      Sequence 1683 BP; 383 A; 387 C; 358 G; 548 T; 7 other;
SQ      Query Match 60.0%; Score 15; DB 16; Length 1683;
        Best Local Similarity 100.0%; Pred. No. 34;
        Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 gttgctgacgtcgt 15
        |||||||
Db      601 GTTGCTGATCTGCT 587

RESULT 9
AAZ23024/c
ID      AAZ23024 standard; cDNA; 1689 BP.
XX
AC      AAZ23024;
XX
DT      17-JAN-2000 (first entry)
XX
DE      Rat kd312 polypeptide encoding cDNA.
XX
KW      kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW      heart attack; head trauma; neurodegenerative disease; rat;
KW      Parkinson's disease; Alzheimer's disease; ss.
XX
OS      Rattus sp.
XX
PN      W09950288-A2.
XX
PD      07-OCT-1999.
XX
PF      30-MAR-1999; 99MO-US06993.
XX
PR      31-MAR-1998; 98US-0053374.
XX
PA      (AMGE-) AMGEN INC.
XX
PI      Yen K;
```

```
XX      WPI; 1999-601322/51.
XX      P-PSDB; AAY42694.
XX      kd312 polypeptides useful for treating diseases and disorders
XX      associated with alterations in cell proliferation and cell death
XX      Claim 2; Fig 9; 85pp; English.
XX      The invention provides nucleic acid molecules encoding human and rat
XX      kd312 polypeptides. The kd312 polypeptides can be expressed by standard
XX      recombinant methodology. The kd312 sequences, and the antibodies against
XX      the proteins may be used to treat or diagnose the presence or progression
XX      of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
XX      stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
XX      Parkinson's disease and Alzheimer's disease). The present sequence
XX      represents the rat kd312 cDNA sequence.
XX      Sequence 1689 BP; 417 A; 462 C; 434 G; 376 T; 0 other;
SQ      Query Match 60.0%; Score 15; DB 20; Length 1689;
        Best Local Similarity 100.0%; Pred. No. 34;
        Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      7 ggatcgtcgtttga 21
        |||||||
Db      495 GCATCTGCTGTTTGA 481

RESULT 10
AAC44132/c
ID      AAC44132 standard; DNA; 2022 BP.
XX
AC      AAC44132;
XX
DT      18-OCT-2000 (first entry)
XX
DE      Zea mays DNA fragment SEQ ID NO: 41750.
XX
KW      Hybridisation assay; genetic mapping; gene expression control;
KW      protein identification; signal transduction pathway; metabolic;
KW      pathway; promoter; termination sequence; corn; ss.
XX
OS      Zea mays subsp. mays.
XX
PN      EP1033405-A2.
XX
PD      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0128845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      28-APR-1999; 99US-0130891.
PR      30-APR-1999; 99US-0131449.
PR      30-APR-1999; 99US-0132048.
PR      04-MAY-1999; 99US-0132407.
PR      05-MAY-1999; 99US-0132484.
PR      06-MAY-1999; 99US-0132485.
PR      06-MAY-1999; 99US-0132486.
PR      06-MAY-1999; 99US-0132487.
```

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161933.
 PR 29-OCT-1999; 99US-0162142.

Query Match 60.0%; Score 15; DB 21; Length 2022;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 lctgctgttgaagc 24
 |||||
 Db 1268 TCTGCTGTTGCAAGC 1254

RESULT 11
 AAX04606/c
 ID AAX04606 standard; DNA; 2034 BP.

AC AAX04606;
 XX 07-APR-1999 (first entry)

DE Maize genomic clone MZU14.4 snor cluster sequence.

XX Small nucleolar RNA gene; snor; snor1.4; maize; promoter;
 KW RNA production; RNA stability; small non-translated RNA;
 KW Intergenic region; U49.1; Snor2.4; U14.4; Snor3.4; ds.

OS Zea sp.

PH Key Location/Qualifiers
 FT misc_feature 1..890

FT /note= "a
 FT /note= "upstream region having promoter activity"
 FT 891..983

FT /tag= b
 FT /note= "Snor1.4 coding sequence; no protein given in
 FT Specification"

FT misc_feature 984..1100
 FT /tag= c
 FT /note= "Intergenic region"
 FT 1101..1294

FT /tag= d
 FT /note= "U49.1 coding sequence; no protein given in
 FT Specification"

FT misc_feature 1295..1361
 FT /tag= e
 FT /note= "Intergenic region"
 FT 1362..1512

FT /tag= f
 FT /note= "Snor2.4 coding sequence; no protein given in
 FT Specification"

FT misc_feature 1513..1587
 FT /tag= g
 FT /note= "Intergenic region"
 FT 1588..1711

FT misc_feature 1712..1743
 FT /tag= h
 FT /note= "U14.4 coding sequence; no protein given in
 FT Specification"

FT misc_feature 1744..1873
 FT /tag= i
 FT /note= "Intergenic region"

FT misc_feature 1874..1973
 FT /tag= j

FT misc_feature 1974..2034
 FT /tag= k

FT /note= "Snor3.1 coding sequence; no protein given in
 FT Specification"
 FT misc_feature 1874..2034
 FT /tag= k
 FT /note= "3' flanking sequence"

PN EP887405-A1.
 PD 30-DEC-1998.

XX 25-JUN-1997; 97EP-0401480.

XX 25-JUN-1997; 97EP-0401480.

XX (GENE-) GENE SHEARS PTY LTD.

XX Brown JWS, Leader DJ;

XX WPI; 1999-047872/05.

XX Producing stabilised RNA molecules - using a DNA precursor
 XX comprising plant small nucleolar RNA stabilising sequences

XX Claim 46; Fig 18a-f; 98pp; English.

XX The present sequence represents the full sequence of the small
 CC nucleolar RNA (snorRNA) cluster of maize genomic clone MZU14.4. The
 CC Snor genes in the cluster are Snor1.4, U49.1, Snor2.4, U14.4, and
 CC Snor3.1. Plant snor genes enable production of multiple stabilised
 CC RNA from a single promoter in a splicing-independent manner,
 CC and allow the production of RNA from intron or non-intron sequences.
 CC The present sequence is used in the method of the invention. The
 CC specification describes a method for producing small non-translated
 CC RNA molecules, which comprises introducing a precursor DNA containing
 CC one or a cluster of coding sequences into a plant or yeast cell. Each
 CC coding sequence comprises a non-translated RNA sequence to be stabilised,
 CC which is capable of functionally interacting with a cellular component
 CC and stabilising sequences comprising at least one protein binding site.
 CC The stabilised RNA units are transcribed by the cell's endogenous
 CC processing machinery. The process is useful for qualitatively and
 CC quantitatively enhancing the production and stability of RNA molecules,
 CC especially non-translated RNA e.g. ribozymes and antisense RNA. The
 CC nucleic acid molecules are useful as probes or primers for identification
 CC of snorRNA sequences and intergenic regions.

XX Sequence 2034 BP; 449 A; 504 C; 435 G; 646 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 2034;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgagctgct 15
 |||||
 Db 971 GTTGTGATCTGCT 957

RESULT 12
 AA23025/c
 ID AA23025 standard; DNA; 3079 BP.

XX AA23025;

XX 17-JAN-2000 (first entry)

XX Rat kd312 genomic DNA sequence.

XX kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
 KW heart attack; head trauma; neurodegenerative disease; rat;
 KW Parkinson's disease; Alzheimer's disease; ss.

XX Rattus sp.

XX

PN W09950288-A2.
 XX 07-OCT-1999.
 PD 30-MAR-1999; 99WO-US06993.
 XX 31-MAR-1998; 98US-0053374.
 PR (AMGE-) AMGEN INC.
 PA Yen K;
 PI WPI: 1999-601322/51.
 DR P-PSDB: AAY42694.
 XX kdl12 polypeptides useful for treating diseases and disorders
 PT associated with alterations in cell proliferation and cell death
 XX
 PS Claim 2; Fig 7; 85pp; English.
 CC The invention provides nucleic acid molecules encoding human and rat
 CC kdl12 polypeptides. The kdl12 polypeptides can be expressed by standard
 CC recombinant methodology. The kdl12 sequences, and the antibodies against
 CC the proteins may be used to treat or diagnose the presence or progression
 CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
 CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
 CC Parkinson's disease and Alzheimer's disease). The present sequence
 CC represents the rat kdl12 genomic DNA sequence.
 XX
 SQ Sequence 3079 BP; 708 A; 843 C; 829 G; 699 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 3079;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 ggaatgctgttga 21
 ||||||||||||
 Db 803 GCATCTGCTTTGA 789

RESULT 13
 AAV49536
 ID AAV49536 standard; DNA: 6703 BP.
 XX
 AC AAV49536;
 XX
 DT 16-NOV-1998 (first entry)
 XX
 DE Adenylate cyclase gene.
 XX
 KW ds; adenylate cyclase; recombinant microorganism; chemical production;
 KM riboflavin; carotenoid.
 XX
 OS Ashbya gossypii.
 XX
 FH Key Location/Qualifiers
 FT 5'UTR 1..670
 FT CDS /*tag= a
 FT 671..6295
 FT /*tag= b
 FT /product= "Adenylate cyclase"
 FT 6296..6703
 FT /*tag= c
 XX
 PN W09829538-A2.
 XX
 PD 09-JUL-1998.
 XX
 PF 29-DEC-1997; 97WO-EP07309.
 XX
 PR 31-DEC-1996; 97CH-0000016.
 XX

PA (BADI) BASF AG.
 XX
 PI Althoeffer H, Altmann-Joehl R, Philippsen P, Seublerger H;
 XX
 DR WPI: 1998-388120/33.
 DR P-PSDB: AAW64518.
 XX
 PT New gene for adenylate cyclase from Ashbya gossypii - useful for
 PT generating recombinant microorganisms with alteration in gene of
 PT cAMP-dependent signalling pathway for increasing production of fine
 PT chemicals
 XX
 PS Claim 1; Page 10-20; 33pp; German.
 XX
 CC The adenylate cyclase (AC) gene is used to generate recombinant
 CC microorganisms for production of chemicals, particularly riboflavin but
 CC also others useful in human or animal nutrition, cosmetics and medicines,
 CC and also carotenoids for use as natural colourings. The modified
 CC microorganisms produce chemicals at higher levels than wild-type
 CC organisms. Particularly, for riboflavin, cells with a disrupted AC gene
 CC are grown on glucose-containing medium.
 XX
 SQ Sequence 6703 BP; 1954 A; 1527 C; 1541 G; 1681 T; 0 other;

Query Match 60.0%; Score 15; DB 19; Length 6703;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 tgctgactgctgt 17
 ||||||||||||
 Db 6234 tgctgactgctgt 6248

RESULT 14
 AA212718
 ID AA212718 standard; CDNA: 275 BP.
 XX
 AC AA212718;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SF0 ID NO:187.
 XX
 KW Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN W09938972-A2.
 PN
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Cirkenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kila D, Labat I;
 PI Lamson G, Lesnikowicz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 DR WPI: 1999-494092/41.
 XX

PT Novel human genes and their expression products which are
PT differentially expressed in different cell types

PS Claim 1; Page 705; 2479pp; English.

XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AA12532 to AA17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AA12532 to AA17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.

XX Sequence 275 BP; 65 A; 44 C; 48 G; 66 T; 52 other;

Query Match 56.0%; Score 14; DB 20; Length 275;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ttgctgagctgct 15
|||||

Db 169 ttgctgagctgct 182

RESULT 15
AA65487/c

ID AAF65487 standard; cDNA: 336 BP.

XX AAF65487;

AC 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 1243.

XX Human; cytosolic; gene therapy: colon cancer; prostate cancer;

KW breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

OS WO200102568-A2.

PN 11-JAN-2001.

PD 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

XX (CHTR) CHIRON CORP.

PA (HSE-) HSE INC.

XX Williams JT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkencjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LM, Strache-Grain B;

XX WPI, 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 722; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.

XX Sequence 336 BP; 97 A; 84 C; 59 G; 96 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 336;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atctgctgttga 22
|||||

Db 159 ATCTGCTGTTGAA 146

RESULT 16
AAV87319/c

ID AAV87319 standard; cDNA: 379 BP.

XX AAV87319;

AC 27-APR-1999 (first entry)

DE EST clone BP797.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;

KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;

KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;

KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

PN WO9845435-A2.

PD 15-OCT-1998.

PF 10-APR-1998; 98WO-US06954.

PR 10-APR-1997; 97US-0835913.

XX (GEMY) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Racie LA, Spaulding V, Treacy M;

XX WPI; 1999-070076/06.

PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries

PS Claim 1; Page 527; 633pp; English.

XX This sequence represents an expressed sequence tag (EST), and is a

Db 350 TCGATCTGCTGTT 337

RESULT 19

AAH10789 standard; cDNA; 393 BP.

AAH10789;

26-JUN-2001 (first entry)

Human cDNA clone (3'-primer) SEQ ID NO:7624.

Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES. INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI: 2001-318749/34.

Claim 3; SEQ ID 7624; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dt primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 393 BP; 89 A; 100 C; 115 G; 82 T; 7 other;

Query Match 56.0%; Score 14; DR 22; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctgctgttgaag 23
|||||

Db 293 tctgctgttgaag 306

RESULT 20

AAI10569 standard; DNA; 464 BP.

AAI10569;

12-OCT-2001 (first entry)

Probe #502 for gene expression analysis in human cervical cell sample.

Probe: human; microarray; gene expression; cervical epithelial cell;

cervical cancer; ss.

Homo sapiens.

WO200157278-A2.

30-JAN-2001; 2001WO-US00670.

09-AUG-2001.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-488901/53.

Claim 25; SEQ ID No 502; 487bp; English.

The present invention relates to human single exon nucleic acid probes (SENPs). The present sequence is one such probe. The SENPs are derived from human HeLa cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 464 BP; 77 A; 137 C; 140 G; 110 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 464;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gctggtctgctgt 17
|||||

Db 165 gctggtctgctgt 178

RESULT 21

AAI131822 standard; DNA; 464 BP.

AAI131822

```

AC  AA131822;
XX
DT  17-OCT-2001 (first entry)
XX
DE  Probe #508 used to measure gene expression in human placenta sample.
XX
KW  Probe: microarray; human; placenta; antenatal diagnosis;
XX  genetic disorder; ss.
XX
OS  Homo sapiens.
XX
PN  WO200157272-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00663.
XX
PR  04-FEB-2000; 2000US-0180312.
XX  26-MAY-2000; 2000US-0207456.
XX  30-JUN-2000; 2000US-0608408.
XX  03-AUG-2000; 2000US-0632366.
XX  21-SEP-2000; 2000US-0234687.
XX  27-SEP-2000; 2000US-0236359.
XX  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI; 2001-488897/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
XX  analyzing gene expression in human placenta -
XX
PS  Claim 25; SEQ ID No 508; 654bp; English.
XX
CC  The present invention relates to single exon nucleic acid probes (SENP).
XX  The present sequence is one such probe. The probes are useful for
XX  producing a microarray for predicting, measuring and displaying gene
XX  expression in samples derived from human placenta. The probes are useful
XX  for antenatal diagnosis of human genetic disorders.
XX
SQ  Sequence 464 BP; 77 A; 137 C; 140 G; 110 T; 0 other;

Query Match          56.0%; Score 14; DB 22; Length 464;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  4 gctgagctcgtctgt 17
    |||||
DB  165 gctgagctcgtctgt 178

RESULT 22
AA100501
ID  AA100501 standard; DNA: 464 BP.
XX
AC  AA100501;
XX
DT  09-OCT-2001 (first entry)
XX
DE  Probe #492 used to measure gene expression in human breast sample.
XX
KW  Probe: human; breast disease; breast cancer; development disorder; ss;
XX  inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS  Homo sapiens.
XX
PN  WO200157270-A2.
XX
PD  09-AUG-2001.
XX

```

```

PF  29-JAN-2001; 2001WO-US00661.
XX
PR  04-FEB-2000; 2000US-0180312.
XX  26-MAY-2000; 2000US-0207456.
XX  30-JUN-2000; 2000US-0608408.
XX  03-AUG-2000; 2000US-0632366.
XX  21-SEP-2000; 2000US-0234687.
XX  27-SEP-2000; 2000US-0236359.
XX  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI; 2001-476286/51.
XX
PT  Novel single exon nucleic acid probe used to measuring gene expression
XX  in a human breast -
XX
PS  Claim 25; SEQ ID No 492; 322bp; English.
XX
CC  The present invention relates to novel single exon nucleic acid probes.
XX  The present sequence is one such probe. The probes are useful for
XX  measuring human gene expression in a human breast sample, where the probe
XX  hybridises at high stringency to a nucleic acid expressed in the human
XX  breast. The probes are useful for predicting, diagnosing, grading,
XX  staging, monitoring and prognosing diseases of the human breast,
XX  particularly those diseases with polygenic aetiology. The diseases
XX  include: breast cancer, disorders of development, inflammatory diseases
XX  of the breast, fibrocystic changes, proliferative breast disease and
XX  non-carcinoma tumours.
XX
CC  Note: The sequence data for this patent did not form part of the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ  Sequence 464 BP; 77 A; 137 C; 140 G; 110 T; 0 other;

Query Match          56.0%; Score 14; DB 22; Length 464;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  4 gctgagctcgtctgt 17
    |||||
DB  165 gctgagctcgtctgt 178

RESULT 23
AA112900
ID  AA112900 standard; DNA: 469 BP.
XX
AC  AA112900;
XX
DT  12-OCT-2001 (first entry)
XX
DE  Probe #2833 for gene expression analysis in human cervical cell sample.
XX
KW  Probe: human; microarray; gene expression; cervical epithelial cell;
XX  cervical cancer; ss.
XX
OS  Homo sapiens.
XX
PN  WO200157278-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00670.
XX
PR  04-FEB-2000; 2000US-0180312.
XX  26-MAY-2000; 2000US-0207456.
XX  30-JUN-2000; 2000US-0608408.
XX  03-AUG-2000; 2000US-0632366.
XX  21-SEP-2000; 2000US-0234687.
XX

```

XX	PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	PN	WPI; 2001-488901/53.
XX	DR	Human genome-derived single exon nucleic acid probes useful for
PT	PF	analyzing gene expression in human cervical epithelial cells -
XX	PS	Claim 25; SEQ ID No 2833; 487bp; English.
XX	CC	The present invention relates to human single exon nucleic acid probes
XX	CC	(SENPs). The present sequence is one such probe. The SENPs are derived
XX	CC	from human HeLa cells. The SENPs can be used to produce a single exon
XX	CC	microarray, which can be used for measuring human gene expression in a
XX	CC	sample derived from human cervical epithelial cells. By measuring gene
XX	CC	expression, the probes are therefore useful in grading and/or staging
XX	CC	of diseases of the cervix, notably cervical cancer.
XX	CC	Note: The sequence data for this patent did not form part of the printed
XX	CC	specification, but was obtained in electronic format directly from WIPO
XX	CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	SQ	Sequence 469 BP; 81 A; 90 C; 145 G; 153 T; 0 other:
XX	Query Match	56.0%; Score 14; DB 22; Length 469;
XX	Best Local Similarity	100.0%; Pred. No. 1.1e+02;
XX	Matches 14: Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	8 gactcgtcgttga 21 	
DB	295 gatctgctgtttga 308	
RESULT 24		
AAI34258		
ID	AAI34258 standard; DNA; 469 BP.	
AC	AAI34258;	
DT	17-OCT-2001 (first entry)	
DE	Probe #2944 used to measure gene expression in human placenta sample.	
KM	Probe: microarray; human; placenta; antenatal diagnosis;	
OS	genetic disorder; ss.	
WO200157272-A2.	Homo sapiens.	
09-AUG-2001.		
30-JAN-2001; 2001WO-US00663.		
04-FEB-2000; 2000US-0180312.		
26-MAY-2000; 2000US-0207456.		
30-JUN-2000; 2000US-0608408.		
23-AUG-2000; 2000US-0632366.		
21-SEP-2000; 2000US-0234687.		
27-SEP-2000; 2000US-0236359.		
04-OCT-2000; 2000GB-0024263.		
(MOLE-) MOLECULAR DYNAMICS INC.		
Penn SG, Hanzel DK, Chen W, Rank DR;		
WPI; 2001-48897/53.		
Human genome-derived single exon nucleic acid probes useful for		

PT	analyzing gene expression in human placenta -
XX	
ES	Claim 25; SEQ ID No 2944; 654pp; English.
XX	
CC	The present invention relates to single exon nucleic acid probes (SENP).
CC	The present sequence is one such probe. The probes are useful for
CC	producing a microarray for predicting, measuring and displaying gene
CC	expression in samples derived from human placenta. The probes are useful
CC	for antenatal diagnosis of human genetic disorders.
XX	
SQ	Sequence 469 BP; 81 A; 90 C; 145 G; 153 T; 0 other;
OY	
Db	Query Match 56.0%; Score 14; DB 22; Length 469; Best Local Similarity 100.0%; Pred. No. 1.1e+02; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
	8 gatctgcgttttga 21 295 gatcgcgttttga 308
RESULT 25	
AAI02819	
ID	AAI02819 standard; DNA; 469 BP.
XX	
AC	AAI02819;
XX	
DT	09-OCT-2001 (first entry)
XX	
DE	Probe #2810 used to measure gene expression in human breast sample.
KW	Probe; human; breast disease; breast cancer; development disorder; ss;
KW	Inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX	
OS	Homo sapiens.
XX	
PN	WO200157270-A2.
PD	
PD	09-AUG-2001.
XX	
PF	29-JAN-2001; 2001MO-US00661.
XX	
PR	04-FEB-2000; 2000US-0180312-
PR	26-MAY-2000; 2000US-0207456-
PR	30-JUN-2000; 2000US-0608408-
PR	03-AUG-2000; 2000US-0632366-
PR	21-SEP-2000; 2000US-0234687-
PR	27-SEP-2000; 2000US-0236359-
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-476286/51.
XX	
PT	Novel single exon nucleic acid probe used to measuring gene expression
XX	in a human breast -
PS	
XX	Claim 25; SEQ ID NO 2810; 322pp; English.
XX	
CC	The present invention relates to novel single exon nucleic acid probes.
CC	The present sequence is one such probe. The probes are useful for
CC	measuring human gene expression in a human breast sample, where the probe
CC	hybridises at high stringency to a nucleic acid expressed in the human
CC	breast. The probes are useful for predicting, diagnosing, grading,
CC	staging, monitoring and prognosing diseases of the human breast,
CC	particularly those diseases with polygenic aetiology. The diseases
CC	include: breast cancer, disorders of development, inflammatory diseases
CC	of the breast, fibrocystic changes, proliferative breast disease and
CC	non-carcinoma tumours.
CC	Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 469 BP; 81 A; 90 C; 145 G; 153 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 476;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatctgctgttga 21
 |||||
 Db 295 gatctgctgttga 308

RESULT 26

AA122139
 ID AA122139 standard; DNA; 476 BP.

AC AA122139;

DT 12-OCT-2001 (first entry)

DE Probe #12072 for gene expression analysis in human cervical cell sample.

KW Probe: human; microarray; gene expression; cervical epithelial cell;
 cervical cancer; ss.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID No 12072; 487pp; English.

PS The present invention relates to human single exon nucleic acid probes

CC (SENP). The present sequence is one such probe. The SENPs are derived

CC from human beta cells. The SENPs can be used to produce a single exon

CC microarray, which can be used for measuring human gene expression in a

CC sample derived from human cervical epithelial cells. By measuring gene

CC expression, the probes are therefore useful in grading and/or staging

CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 476;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatctgctgttga 21
 |||||
 Db 135 gatctgctgttga 148

RESULT 27

AA147435
 ID AA147435 standard; DNA; 476 BP.

AC AA147435;

DT 17-OCT-2001 (first entry)

DE Probe #16121 used to measure gene expression in human placenta sample.

KW Probe: microarray; human; placenta; antenatal diagnosis;
 genetic disorder; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -

XX Claim 25; SEQ ID No 16121; 654pp; English.

PS The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders.

XX Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 476;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatctgctgttga 21
 |||||
 Db 135 gatctgctgttga 148

RESULT 28

AA107839
 ID AA107839 standard; DNA; 476 BP.

AC AA107839;

DT 09-OCT-2001 (first entry)

DE Probe #7830 used to measure gene expression in human breast sample.

Probe: human: breast disease; breast cancer; development disorder; ss;
 inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI: 2001-476286/51.
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 XX in a human breast -
 PS
 SS Claim 25; SEQ ID No 7830; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX SQ Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other:
 XX
 XX
 Query Match 56.0%; Score 14; DB 22; Length 476;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 gatcgtcgttga 21
 |||||
 Db 135 gatcgtcgttga 148
 RESULT 29
 AAI19821
 ID AAI19821 standard; DNA: 558 BP.
 XX
 AC AAI19821;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Probe #9754 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe: human: microarray; gene expression; cervical epithelial cell;
 KM cervical cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX

PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI: 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human cervical epithelial cells -
 PS
 SS Claim 25; SEQ ID No 9754; 487pp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX SQ Sequence 558 BP; 95 A; 154 C; 169 G; 140 T; 0 other:
 XX
 XX
 Query Match 56.0%; Score 14; DB 22; Length 558;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 gctgacgtcgtt 17
 |||||
 Db 105 gctgacgtcgtt 118
 RESULT 30
 AAI45016
 ID AAI45016 standard; DNA: 558 BP.
 XX
 AC AAI45016;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #13702 used to measure gene expression in human placenta sample.
 XX
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KM genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR

PR 04-OCT-2000: 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID NO 13702; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENPs).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 558 BP; 95 A; 154 C; 169 G; 140 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 558;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gctgagctcgtctgt 17
 |||||||
 Db 105 gctgagctcgtctgt 118

RESULT 31
 AAI05537
 ID AAI05537 standard; DNA; 558 BP.
 XX
 AC AAI05537;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Probe #5528 used to measure gene expression in human breast sample.
 XX
 KW Probe: human; breast disease; breast cancer; development disorder; ss;
 KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001: 2001WO-US00661.
 XX
 PR 04-FEB-2000: 2000US-0180312.
 XX
 PR 26-MAY-2000: 2000US-0207456.
 PR 30-JUN-2000: 2000US-0608408.
 PR 03-AUG-2000: 2000US-0632366.
 PR 21-SEP-2000: 2000US-0234687.
 PR 27-SEP-2000: 2000US-0236359.
 PR 04-OCT-2000: 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-476286/51.
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 XX in a human breast -
 XX
 PS Claim 25; SEQ ID No 5528; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for

CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 558 BP; 95 A; 154 C; 169 G; 140 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 558;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gctgagctcgtctgt 17
 |||||||
 Db 105 gctgagctcgtctgt 118

RESULT 32
 AAI18864
 ID AAI18864 standard; DNA; 575 BP.
 XX
 AC AAI18864;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Probe #8797 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe: human; microarray; gene expression; cervical epithelial cell;
 KM cervical cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001: 2001WO-US00670.
 XX
 PR 04-FEB-2000: 2000US-0180312.
 PR 26-MAY-2000: 2000US-0207456.
 PR 30-JUN-2000: 2000US-0608408.
 PR 03-AUG-2000: 2000US-0632366.
 PR 21-SEP-2000: 2000US-0234687.
 PR 27-SEP-2000: 2000US-0236359.
 PR 04-OCT-2000: 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 25; SEQ ID No 8797; 487pp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at fip.wipo.int/pub/published_pcl_sequences.
 XX
 SO Sequence 575 BP; 174 A; 128 C; 161 G; 112 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 575;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tggatctgctgttt 19
 |||||||
 DB 222 tggatctgctgttt 235

RESULT 33
 AAC99862
 ID AAC99862 standard; CDNA; 612 BP.

AC AAC99862;

DT 17-OCT-2001 (first entry)

DE Probe #12672 used to measure gene expression in human placenta sample.

XX Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -

XX Claim 25; SEQ ID NO 12672; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).

CC The present invention is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders.

SO Sequence 575 BP; 174 A; 128 C; 161 G; 112 T; 0 other;

AAC99862
 ID AAC99862 standard; CDNA; 612 BP.
 XX
 AC AAC99862;
 DT 13-MAR-2001 (first entry)
 DE Human secreted protein gene 45 SEQ ID NO:55.
 XX
 XX Human: secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytoskeletal; cardiac; vasotropic;
 KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; gene therapy; pathological condition;
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischemia;
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; skin aging; food additive; preservative; ss.
 XX
 OS Homo sapiens.
 PN WO200070042-A1.
 PD 23-NOV-2000.
 PF 11-MAY-2000; 2000WO-US12788.
 PR 13-MAY-1999; 99US-0134068.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
 PI Duan RD, Florence KA, Soppet DR;
 DR WPI; 2000-679828/66.
 DR P-PSDB; AAB56121.
 XX Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX
 PS Claim 1; Page 881; 1065pp; English.
 XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
 CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
 CC proteins have activities based on the tissues and cells the genes are
 CC expressed in. Examples of activities include: immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytoskeletal; cardiac;
 CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
 CC virucide; fungicide; and ophthalmological. The human secreted
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological
 CC condition or susceptibility to a pathological condition. Disorders which
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
 CC disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
 CC fungi and ocular disorders e.g. corneal infection. The proteins can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The proteins can also be used as a
 CC food additive or preservative to increase or decrease storage
 CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
 CC in the exemplification of the present invention.

SO Sequence 612 BP; 199 A; 139 C; 139 G; 135 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 612;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 34

OY 2 ttctgagatctgct 15
 DB 372 ttctgagatctgct 385

RESULT 35
 ID AAA44124/c
 AC AAA44124 standard; cDNA; 619 BP.
 XX
 DT 21-AUG-2000 (first entry)
 DE
 XX Human secreted expressed sequence tag SEQ ID NO:699.
 KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
 KW antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic;
 KW cerebroprotective; anticoagulant; antidepressant; gene therapy;
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.
 KW
 OS Homo sapiens.
 PN WO200021991-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-US24206.
 XX
 PR 15-OCT-1998; 98US-0104436.
 XX
 PA (GENV) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
 PI Metberg D, Treacy M, Bowman MR;
 DR
 XX WPI: 2000-317938/27.
 PT Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders -
 PT such as autoimmune, infectious, and central nervous system disorders -
 PS
 XX Claim 1; Page 383; 803pp; English.
 CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
 CC antiasthmatic; vulnerary; antitumor; osteoprotective; neuroprotective;
 CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
 CC anticoagulant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation

CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
 CC in the exemplification of the present invention.
 CC
 SQ Sequence 619 BP; 193 A; 123 C; 125 G; 178 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 619;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 attctgcttttgaa 22
 DB 43 ATCTGCTGTGGAA 30

RESULT 36
 ID AAF14018/c
 AC AAF14018 standard; cDNA; 707 BP.
 XX
 AC AAF14018;
 XX
 DT 13-MAR-2001 (first entry)
 DE
 XX Aspergillus oryzae EST SEQ ID NO:6541.
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 KW
 OS Aspergillus oryzae.
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 DR
 XX WPI: 2000-594572/56.
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 PS
 XX Claim 88; Page 2678; 3161pp; English.
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an

CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

CC Sequence 707 BP: 154 A; 241 C; 156 G; 156 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 707;

Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Gaps 0;

Matches 14; Conservative 0; Indels 0; Gaps 0;

OY 5 ctgagatctgctg 18
|||||

DB 693 CTGATCTGCTCTT 680

RESULT 37

AA092502
ID AA092502 standard; cDNA: 735 BP.

AC AA092502;

DT 07-FEB-1996 (first entry)

DE Mouse antibody F4-7 heavy chain variable region coding sequence.

XX Primer: amplification; PCR; mouse; kappa chain; heavy chain; Fab;
XX antibody; immunoreactive; animal; variegated display library;
XX variable region; antigen; immunorecessive; cell surface marker; foetal;
XX cancer; stem cell; variant; therapy; Alzheimer's disease; hybridoma;
XX familial hypercholesterolaemia; binding affinity; ds.

OS Mus musculus.

XX Key Location/Qualifiers

FT CDS 67..735
FT /tag= a
FT /product= antibody F4-7 heavy chain variable region

PN W09515982-A2.

XX 15-JUN-1995.

PF 08-DEC-1994; 94WO-US14106.

PR 06-DEC-1994; 94US-0350400.

PR 08-DEC-1993; 93US-0164022.

XX (GEN2) GENZYME CORP.

PI Barsomian G, Copeland DP, Hillhouse D, Johnson T;

DR WPI: 1995-224291/29.

DR P-PSDB: AAR5458.

PT Generating new antibodies specific for immunorecessive epitopes
PT by selection from variegated V gene library cloned from
PT immuno:tolerance derived antibody repertoire, useful in diagnosis,
PT purificn. and therapy, e.g. of cancer

XX Disclosure: Page 77-78; 109pp; English.

XX The coding sequence of the heavy chain variable region from the mouse
CC antibody F4-7. This sequence was isolated from a variegated display
CC library (VDL) of variable regions derived from a repertoire of
CC antibodies from an immunotolerised animal. The VDL is generated by PCR
CC amplifying the variable regions from the antibody coding sequences using
CC the primers AAO74153-74. The variable regions, esp the complementarity
CC determining regions (CDR; see AAR5462-93 for examples of CDRs) from the

CC immunotolerant animals' antibodies are used to construct an antibody
CC against a immunorecessive antigen e.g. a cell surface marker on a foetal,
CC cancer or stem cell, which can differentiate between variant or related
CC forms of the antigen. The antibodies generated can be used in the
CC diagnosis, e.g. detection of the immunorecessive antigen, or in therapy
CC e.g. of cancer, Alzheimer's disease or familial hypercholesterolaemia.
CC The method of production of the antibody allows rapid and sensitive
CC isolation of antibodies that would be difficult to isolate by standard
CC methods. The antibodies produced have greater binding affinity than
CC those produced by combinatorial/hybridoma methods.

CC Sequence 735 BP: 172 A; 213 C; 188 G; 162 T; 0 other;

Query Match 56.0%; Score 14; DB 16; Length 735;

Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Gaps 0;

Matches 14; Conservative 0; Indels 0; Gaps 0;

OY 3 tgcgtgactgctg 16
|||||

DB 465 tgcgtgactgctg 478

RESULT 38

AA200347/C
ID AA200347 standard; DNA: 753 BP.

AC AA200347;

DT 26-OCT-1999 (first entry)

DE Nucleotide sequence of human hsfATP3.

XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
XX fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.

XX Homo sapiens.

XX W09936537-A2.

XX 22-JUL-1999.

PF 14-JAN-1999; 99WO-US00182.

PR 14-JAN-1999; 99US-0232201.

PR 15-JAN-1998; 98US-0071374.

PR 20-JUL-1998; 98US-0093491.

PR 04-DEC-1998; 98US-0110941.

PR 14-JAN-1999; 99US-0232195.

PR 14-JAN-1999; 99US-0232197.

XX (MILL-) MILLENNIUM PHARM INC.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

DR WPI: 1999-444398/37.

DR P-PSDB: AAV14937.

PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease

XX Example 1; Fig 16; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid

CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCPA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.

XX Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 753;
 Best local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctgctgttgaag 23
 Db 436 TCTGCTGTTCGAG 423

RESULT 39
 AAF89005/c
 ID AAF89005 standard; DNA: 753 BP.

XX AAF89005;

DT 06-JUL-2001 (first entry)

XX Human FATP3 coding sequence SEQ ID NO: 14.

KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
 KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
 KW weight control; tuberculosis; TB; anti-fungal; ds.

OS Homo sapiens.

PN WO200121795-A2.

PD 29-MAR-2001.

PF 21-SEP-2000; 2000WO-US25891.

PR 23-SEP-1999; 99US-0405504.

PR 23-SEP-1999; 99US-0405505.

PR 16-DEC-1999; 99US-0465280.

PR 17-FEB-2000; 2000US-0506252.

PR 06-JUL-2000; 2000US-0611197.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

PA (MILL-) MILLENNIUM PHARM INC.

PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;

DR WPI; 2001-354783/37.

XX P-PSDB; AAB83227.

PT New fatty acid transport proteins (FATPs) useful for the manufacture of

XX medicament for treating obesity, diabetes and heart disease -

PS Disclosure: Fig 16; 287pp; English.

XX The present invention provides the protein and coding sequences of fatty
 CC acid transport proteins (FATPs) from a number of species, including
 CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
 CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus
 CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
 CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
 CC treat TB. That from M. grisea (also known as rice blast fungus) can be
 CC used to develop anti-fungal agents capable of preventing infection of
 CC rice. Those from the human can be used to develop treatments for

CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
 CC present sequence is one of the sequences described in the exemplification
 CC of the invention.

XX Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 753;
 Best local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctgctgttgaag 23
 Db 436 TCTGCTGTTCGAG 423

RESULT 40
 AAF15012/c
 ID AAF15012 standard; CDNA: 769 BP.

XX AAF15012;

DT 13-MAR-2001 (first entry)

XX Trichoderma reesei EST SEQ ID NO: 7535.

XX Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

KW metabolic pathway engineering; catabolic pathway engineering; ss.

OS Trichoderma reesei.

PN WO200056762-A2.

PD 28-SEP-2000.

PF 22-MAR-2000; 2000WO-US07781.

PR 22-MAR-1999; 99US-0273623.

XX (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

DR WPI; 2000-594572/56.

XX Claim 89; Page 3043-3044; 3161pp; English.

XX The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the

CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fuguarum venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 CC
 XX
 SQ Sequence 769 BP: 203 A; 234 C; 203 G; 114 T; 15 other:
 Query Match 56.0%; Score 14; DB 21; Length 769;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 3 tgcgtgacgtcgtg 16
 ||||||||||||
 DB 284 TGCTGATCTGCTG 271
 RESULT 41
 AAZ06233
 ID AAZ06233 standard; DNA: 891 BP.
 XX
 AC AAZ06233;
 DT 30-SEP-1999 (first entry)
 DE
 XX
 XX Human secreted protein gene No. 15.
 XX
 KW Human: secreted protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 PN WO935158-A1.
 XX
 PD 15-JUL-1999.
 XX
 PF 06-JAN-1999; 99WO-US00108.
 XX
 PR 07-JAN-1998; 98US-0070704.
 PR 07-JAN-1998; 98US-0070657.
 PR 07-JAN-1998; 98US-0070658.
 PR 07-JAN-1998; 98US-0070692.
 XX
 PA (HOMA-) HUMAN GENOME SCI INC.
 PI
 PI Brewer LA, Duan RD, Edner R, Lafleur DW, NJ J;
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
 DR WPI: 1999-444190/37.
 DR P-PSDB: AAY38400, AAY38453.
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 XX
 PS Claim 1; Page 158; 227pp; English.
 XX
 CC This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number is given in the descriptor line.
 CC The gene can be used to generate fusion proteins by linking to the gene
 CC to a human immunoglobulin Fc portion (e.g. AAZ06210) for increasing the
 CC stability of the fused protein as compared to the human protein only.
 CC The invention relates to 36 novel genes and their fragments (nucleic
 CC acid sequences: AAZ06219-206263; amino acid sequences AAY38386-138498)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new

CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 36
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAZ06219 for described uses).
 CC
 XX
 SQ Sequence 891 BP: 207 A; 214 C; 220 G; 250 T; 0 other:
 Query Match 56.0%; Score 14; DB 20; Length 891;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2 ttgcgtgacgtcgt 15
 ||||||||||||
 DB 727 ttgcgtgacgtcgt 740
 RESULT 42
 AAX82834/C
 ID AAX82834 standard; DNA: 1044 BP.
 XX
 AC AAX82834;
 DT 30-JUN-2000 (first entry)
 DE
 XX L-asparaginase DNA.
 XX
 KW L-asparaginase; fermentation; ss.
 KW
 XX
 OS Unidentified.
 XX
 PN CN1237633-A.
 XX
 PD 08-DEC-1999.
 XX
 PF 01-JUN-1998; 98CN-0102046.
 XX
 PR 01-JUN-1998; 98CN-0102046.
 XX
 PA (MICR-) INST MICROBIOLOGY CHINESE ACAD SCT.
 PI
 PI Qian S, Wang Y, Meng G;
 PI
 DR WPI: 2000-351194/31.
 DR P-PSDB: AAW90867.
 XX
 PT New structured L-asparaginase bacterial host cell and its production
 PT culture -
 XX
 PS Disclosure; Fig 2; 12pp; Chinese.
 XX
 CC This invention describes a novel L-asparaginase-producing recombinant
 CC bacteria and its fermentation culture method. The new cell is produced
 CC using standard recombinant techniques. The enzyme gene expressed in
 CC the cell is different from the reported enzyme gene in sequence, and
 CC its individual amino acid composition is also different. The culture
 CC medium adopted for culturing the new bacterial cells is cheap and easily
 CC available, and the fermentation process used to grow the cells and
 CC express the enzyme is simple and easy to implement. This sequence
 CC encodes an L-asparaginase which is described in the method of the
 CC invention.
 CC
 XX
 SQ Sequence 1044 BP: 274 A; 281 C; 272 G; 217 T; 0 other:
 Query Match 56.0%; Score 14; DB 21; Length 1044;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 3 tgcgtgacgtcgtg 16
 ||||||||||||
 DB 1028 TGCTGATCTGCTG 1015

```

RESULT 43
AAZ09784
ID AAZ09784 standard; DNA: 1155 BP.
XX
XX
AC AAZ09784:
XX
XX
DT 23-NOV-1999 (first entry)
DE
DE E. coli S-adenosyl methionine synthase SAM DNA.
XX
XX
KM SAM: S-adenosyl methionine synthase; biosynthesis; biotin; bios1; bios2;
KM bios3; cofactor; decarboxylation; Vitamin H; ss.
XX
XX
OS Escherichia coli.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1155
FT /tag= a
FT /product= "S-adenosyl methionine synthase"
XX
XX
PN DE19606872-A1.
XX
XX
PD 26-AUG-1999.
XX
XX
PF 19-FEB-1998: 98DE-1006872.
XX
XX
PR 19-FEB-1998: 98DE-1006872.
XX
XX
PA (BAD1 ) BASF AG.
XX
XX
PI Schroeder H;
XX
XX
DR WPI: 1999-480095/41.
XX
XX
DR P-PSDB; AAY33263.
XX
XX
PT Production of biotin by expressing S-adenosyl-methionine synthase and
PT second biotin synthesis gene in host cells -
XX
XX
PS Claim 1; Page 8-10; 48pp; German.
XX
XX
CC This invention describes a novel method for the preparation of biotin
CC (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC sequence (I), and (b) at least one of the other biotin biosyntheses
CC genes bios1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC most animals and some microorganisms. Expression of biotin plus bios1,
CC bios2 or bios3 significantly increases productivity of biotin
CC biosynthesis, particularly by at least 3 times. This sequence encodes
CC the Escherichia coli SAM protein which is used in the method of the
CC invention.
XX
XX
SQ Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other.

Query Match 56.0%; Score 14; DB 20; Length 1155;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactgcgtg 16
|||||
DB 1031 tgcgtgactgcgtg 1044

RESULT 44
AAA99433
ID AAA99433 standard; DNA: 1155 BP.
XX
XX
AC AAA99433:
XX
XX
DT 26-JAN-2001 (first entry)

```

```

DE Met K oligonucleotide SEQ ID 17 used in preparation of L-methionine.
XX
XX
XX L-methionine production; Met; Escherichia coli; ds.
XX
XX
OS Escherichia coli.
XX
XX
PN JP2000139471-A.
XX
XX
PD 23-MAY-2000.
XX
XX
PF 17-NOV-1998; 98JP-0326717.
XX
XX
PR 17-NOV-1998; 98JP-0326717.
XX
XX
PA (AJIN ) AJINOMOTO KK.
XX
XX
DR WPI: 2001-018703/03.
XX
XX
DR P-PSDB; AAB26814.
XX
XX
PT Fermentative process for preparation of L-methionine (Met), comprises
PT using a modified Met producing microorganism, particularly an
PT Escherichia species
XX
XX
PS Example 2; Page 17-18; 23pp; Japanese.
XX
XX
CC The invention relates to a process for preparing L-methionine (Met)
CC using a modified Met producing organism. The modified microorganism has
CC its Met biosynthetic system repressor deleted, and has enhanced
CC homoserine transsuccinylase activity. The organism optionally has an
CC attenuated intracellular S-adenosyl methionine synthetase (SAM). The
CC process is used for the preparation of L-methionine. Sequences AAA99416
CC and AAA99433 encode proteins represented in AAB26816 and AAB26814
CC respectively, which are involved in the Escherichia coli Met biosynthetic
CC pathway. AAA99417-A99441 (excluding AAA99433) represent primers, all
CC sequences are used in the production of the modified Met producing
CC organism of the invention.
XX
XX
SQ Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other.

Query Match 56.0%; Score 14; DB 22; Length 1155;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactgcgtg 16
|||||
DB 1031 tgcgtgactgcgtg 1044

RESULT 45
AAZ00358/C
ID AAZ00358 standard; DNA: 1173 BP.
XX
XX
AC AAZ00358:
XX
XX
DT 26-OCT-1999 (first entry)
XX
XX
DE Nucleotide sequence of human hSFATP3.
XX
XX
KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
XX
OS Homo sapiens.
XX
XX
PN WO9936537-A2.
XX
XX
PD 22-JUL-1999.
XX
XX
PF 14-JAN-1999; 99WO-US00182.
XX
XX
PR 14-JAN-1999; 99US-0232201.
XX
XX
PR 15-JAN-1998; 98US-0071374.
XX
XX
PR 20-JUL-1998; 98US-0093491.

```

PR 04-DEC-1998; 9805-0110941.
 PR 14-JAN-1999; 9905-0232195.
 PR 14-JAN-1999; 9905-0232197.
 PR 14-JAN-1999; 9905-0232200.

XX
 PA (MILL-) MILLENNIUM PHARM INC.
 PA (WHEED) WHITEHEAD INST BIOMEDICAL RES.

PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA:

DR WPI: 1999-444398/37.

DR P-PSDB: AAY14948.

XX
 PT Fatty acid transport proteins and related polynucleotides, useful
 for treating obesity, diabetes and heart disease

PS Claim 31; Fig 48; 255pp; English.

XX
 CC The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.

XX
 SQ Sequence 1173 BP: 247 A; 327 C; 351 G; 248 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1173;

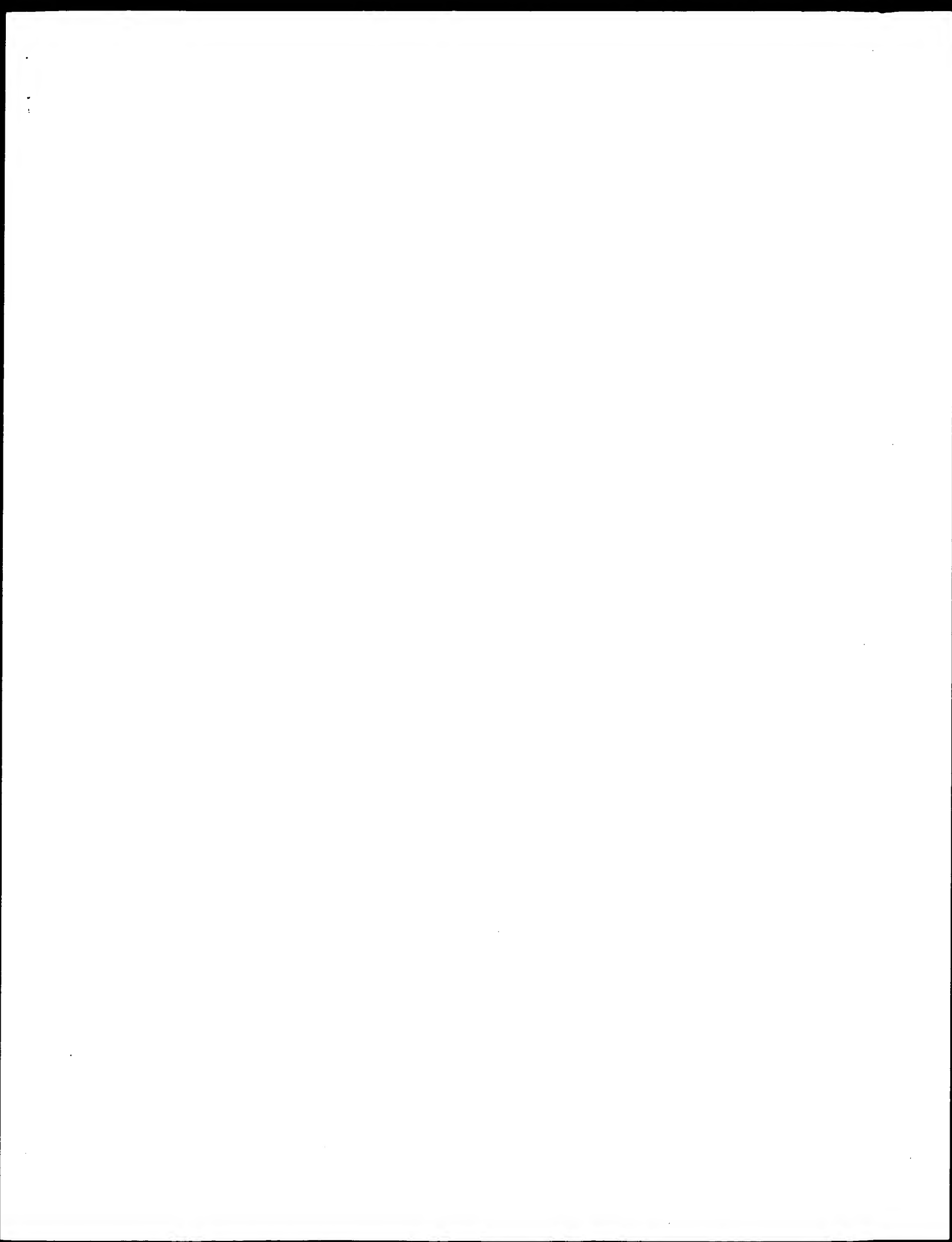
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctgctgttgaag 23
 |||

DB 853 TCTGCTGTTGAAG 840

Search completed: December 26, 2001, 14:16:34
 Job time: 6986 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:52:31 ; Search time 143.1 Seconds
(Without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196f-3

Perfect score: 25

Sequence: 1 gtcctgacatctgcttgaagcg 25

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 14184

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents NA:

1: /cgn2_6/pdata/2/ina/5A.COMB.seq:*
2: /cgn2_6/pdata/2/ina/5B.COMB.seq:*
3: /cgn2_6/pdata/2/ina/6A.COMB.seq:*
4: /cgn2_6/pdata/2/ina/6B.COMB.seq:*
5: /cgn2_6/pdata/2/ina/PCRTUS.COMB.seq:*
6: /cgn2_6/pdata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appl1
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appl1
3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appl1
4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appl1
5	16	64.0	5000	3	US-09-104-070-1	Sequence 1, Appl1
6	15	60.0	708	4	US-08-998-416-997	Sequence 997, App
7	14	56.0	384	2	US-08-820-170A-8	Sequence 8, Appl1
8	14	56.0	384	3	US-09-055-699-8	Sequence 8, Appl1
9	14	56.0	384	4	US-09-273-565-8	Sequence 8, Appl1
10	14	56.0	735	5	PCRT-US94-14106-54	Sequence 54, Appl1
11	14	56.0	753	4	US-09-232-191-14	Sequence 14, Appl1
12	14	56.0	753	4	US-09-232-200-14	Sequence 14, Appl1
13	14	56.0	753	4	US-09-232-197-14	Sequence 14, Appl1
14	14	56.0	1173	4	US-09-232-200-50	Sequence 50, Appl1
15	14	56.0	1173	4	US-09-232-197-6	Sequence 6, Appl1
16	14	56.0	1493	2	US-08-820-170A-9	Sequence 9, Appl1
17	14	56.0	1493	3	US-09-055-699-9	Sequence 9, Appl1
18	14	56.0	1493	4	US-09-273-565-9	Sequence 15, Appl1
19	14	56.0	1848	1	US-08-447-422-15	Sequence 68, Appl1
20	14	56.0	1998	4	US-09-232-200-68	Sequence 68, Appl1
21	14	56.0	1998	4	US-09-232-197-68	Sequence 68, Appl1
22	14	56.0	2087	4	US-09-232-191-6	Sequence 6, Appl1
23	14	56.0	2087	4	US-09-232-200-6	Sequence 6, Appl1
24	14	56.0	2087	4	US-09-232-197-6	Sequence 6, Appl1
25	14	56.0	2166	4	US-09-232-200-101	Sequence 101, App
26	14	56.0	2166	4	US-09-232-197-101	Sequence 101, App
27	14	56.0	3243	2	US-08-239-276-6	Sequence 6, Appl1

C 28	14	56.0	3243	2	US-08-468-579B-6	Sequence 6, Appl1
C 29	14	56.0	3243	3	US-08-468-579B-6	Sequence 6, Appl1
C 30	14	56.0	3282	1	US-08-072-574-11	Sequence 11, Appl1
C 31	14	56.0	3282	1	US-08-486-270-11	Sequence 11, Appl1
C 32	14	56.0	3282	3	US-08-367-264-11	Sequence 11, Appl1
C 33	14	56.0	4085	1	US-08-072-574-7	Sequence 7, Appl1
C 34	14	56.0	4085	1	US-08-486-270-7	Sequence 7, Appl1
C 35	14	56.0	4085	3	US-08-367-264-7	Sequence 7, Appl1
C 36	14	56.0	4181	1	US-08-072-574-9	Sequence 9, Appl1
C 37	14	56.0	4181	1	US-08-486-270-9	Sequence 9, Appl1
C 38	14	56.0	4181	3	US-08-367-264-9	Sequence 9, Appl1
C 39	14	56.0	4207	4	US-08-660-148-3	Sequence 3, Appl1
C 40	14	56.0	4207	4	US-08-660-148-3	Sequence 3, Appl1
C 41	14	56.0	4303	4	US-08-660-148-4	Sequence 4, Appl1
C 42	14	56.0	4303	4	US-08-660-148-6	Sequence 6, Appl1
C 43	14	56.0	4566	2	US-08-465-976A-1	Sequence 1, Appl1
C 44	14	56.0	4566	2	US-08-982-412-1	Sequence 1, Appl1
C 45	14	56.0	5319	1	US-08-169-927-1	Sequence 1, Appl1
C 46	14	56.0	7326	1	US-08-194-468-1	Sequence 1, Appl1
C 47	14	56.0	7344	1	US-08-961-739-1	Sequence 1, Appl1
C 48	14	56.0	12284	2	US-08-876-991-1	Sequence 1, Appl1
C 49	14	56.0	12284	2	US-09-059-853-1	Sequence 1, Appl1
C 50	14	56.0	28804	2	US-08-592-874-1	Sequence 2, Appl1
C 51	14	56.0	28804	3	US-09-096-942-2	Sequence 2, Appl1
C 52	13	52.0	28804	3	US-09-096-867-2	Sequence 2, Appl1
C 53	13	52.0	749	4	US-08-998-416-727	Sequence 727, App
C 54	13	52.0	1161	5	US-08-153-848-31	Sequence 31, Appl1
C 55	13	52.0	1161	5	PCRT-US93-11153-31	Sequence 31, Appl1
C 56	13	52.0	1242	3	US-08-252-966B-13	Sequence 13, Appl1
C 57	13	52.0	1729	3	US-08-188-930-68	Sequence 68, Appl1
C 58	13	52.0	1835	1	US-08-361-611-3	Sequence 3, Appl1
C 59	13	52.0	1835	2	US-08-565-655-3	Sequence 3, Appl1
C 60	13	52.0	1835	2	US-08-946-967-3	Sequence 3, Appl1
C 61	13	52.0	1862	3	US-08-999-723-1	Sequence 1, Appl1
C 62	13	52.0	1862	4	US-09-434-427-1	Sequence 1, Appl1
C 63	13	52.0	2143	4	US-09-064-693A-18	Sequence 18, Appl1
C 64	13	52.0	2254	1	US-08-153-848-27	Sequence 27, Appl1
C 65	13	52.0	2254	5	US-09-299-843A-27	Sequence 27, Appl1
C 66	13	52.0	2254	5	PCRT-US93-11153-27	Sequence 27, Appl1
C 67	13	52.0	2492	4	US-09-381-779-1	Sequence 1, Appl1
C 68	13	52.0	2674	2	US-08-926-774-2	Sequence 2, Appl1
C 69	13	52.0	3003	2	US-08-873-479-41	Sequence 41, Appl1
C 70	13	52.0	3119	3	US-09-299-843A-31	Sequence 31, Appl1
C 71	13	52.0	3206	1	US-08-373-134D-1	Sequence 1, Appl1
C 72	13	52.0	3206	2	US-09-114-637-1	Sequence 1, Appl1
C 73	13	52.0	3435	5	US-08-366-577-1	Sequence 1, Appl1
C 74	13	52.0	3435	5	PCRT-US96-00005-1	Sequence 9, Appl1
C 75	13	52.0	3582	2	US-08-611-729A-9	Sequence 9, Appl1
C 76	13	52.0	3603	1	US-08-188-582-15	Sequence 15, Appl1
C 77	13	52.0	3603	1	US-08-646-715-15	Sequence 15, Appl1
C 78	13	52.0	3731	4	US-08-811-583-1	Sequence 1, Appl1
C 79	13	52.0	3731	4	US-09-106-638-1	Sequence 1, Appl1
C 80	13	52.0	4800	4	US-08-920-827-18	Sequence 18, Appl1
C 81	13	52.0	4856	4	US-09-045-360-1	Sequence 18, Appl1
C 82	13	52.0	4954	1	US-08-920-827-18	Sequence 18, Appl1
C 83	13	52.0	4954	1	US-08-921-177C-18	Sequence 18, Appl1
C 84	13	52.0	4954	1	US-08-362-577C-18	Sequence 18, Appl1
C 85	13	52.0	4954	2	US-08-920-828-18	Sequence 25, Appl1
C 86	13	52.0	4954	2	US-08-064-693A-25	Sequence 25, Appl1
C 87	13	52.0	6641	4	US-08-149-097D-24	Sequence 24, Appl1
C 88	13	52.0	7032	3	US-08-949-366-24	Sequence 24, Appl1
C 89	13	52.0	7032	3	US-08-450-562-24	Sequence 25, Appl1
C 90	13	52.0	7032	3	US-08-949-366-25	Sequence 25, Appl1
C 91	13	52.0	7089	3	US-08-759-444-2	Sequence 2, Appl1
C 92	13	52.0	7089	3	US-08-450-562-25	Sequence 2, Appl1
C 93	13	52.0	8880	1	US-08-680-897-1	Sequence 3, Appl1
C 94	13	52.0	8880	2	US-08-147-777-3	Sequence 3, Appl1
C 95	13	52.0	24979	2	US-08-452-877-3	Sequence 3, Appl1
C 96	13	52.0	24979	5	PCRT-US93-03965-3	Sequence 3, Appl1
C 97	13	52.0	24979	5	US-08-471-810A-2	Sequence 2, Appl1
C 98	13	52.0	4403765	1	US-09-103-840A-2	Sequence 1, Appl1
C 99	13	52.0	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
C 100	13	52.0	4411529	4	US-09-103-840A-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1

US-08-401-068-7

Sequence 7, Application US/08401068

Patent No. 5859335

GENERAL INFORMATION:

APPLICANT: Patton, David

TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/401,068

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/351,970

FILING DATE: 08-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8614

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1041 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1038

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /product= "biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;

Best Local Similarity 100.0%; Pred. No. 2.7e-05;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgagtcgtctgttgaacg 25

|||||

Db 57 GTTCTGATCTCTCTGTTGAACG 81

RESULT 2

US-08-846-338-7

Sequence 7, Application US/08846338

Patent No. 5869719

GENERAL INFORMATION:

APPLICANT: Patton, David

TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/846,338

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Mels, J. Timothy

REGISTRATION NUMBER: 38,241

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1041 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1038

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /product= "biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;

Best Local Similarity 100.0%; Pred. No. 2.7e-05;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgagtcgtctgttgaacg 25

|||||

Db 57 GTTCTGATCTCTCTGTTGAACG 81

RESULT 3

US-08-411-768B-1

Sequence 1, Application US/08411768B

Patent No. 6083712

GENERAL INFORMATION:

APPLICANT: Olwen Birch

APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw

TITLE OF INVENTION: Biotechnological Method

of Producing Biotin

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect

SOFTWARE: Version 5.1

CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
IMMEDIATE SOURCE:
STRAIN: DSM498
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="b10b"
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="b10c"
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="b10a"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: /aminoacid="amino"
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown", involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="b10b RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="b10c RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="b10a RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter plac"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 2.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtgtgtgattcgtcgtttgaagcg 25
Db 173 GTTGTGATCTGCTGTTGAAGCG 197

RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
Version 5.1

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411.768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pBO30A15-9
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1154..2308
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start=1154
; OTHER INFORMATION: /EC_number=2.3.1.47
; OTHER INFORMATION: /product="KAPA synthase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /gene="bioF"
; OTHER INFORMATION: /number=2
; OTHER INFORMATION: /standard_name="8-Amino-7-oxononanoate synthase"
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3043..3753
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start=3043
; OTHER INFORMATION: /EC_number=6.3.3.3
; OTHER INFORMATION: /product="DTB synthase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /gene="bioD"
; OTHER INFORMATION: /number=4
; OTHER INFORMATION: /standard_name="Dethiolobiotin synthase"
;
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 1141..1156
; OTHER INFORMATION: /standard_name="bioF RBS"
;
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3030..3045
; OTHER INFORMATION: /standard_name="bioD RBS"
;
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
;
; US-08-411-768B-6
;
Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 2.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 1 gttgcgagatcgtcgttgaagc 25
|||||
Db 173 gttgcgagatcgtcgttgaagc 197

```

```

; GENERAL INFORMATION:
; APPLICANT: Richards, Eric J.
; TITLE OF INVENTION: DNA Methylation Gene from Plants
; FILE REFERENCE: Wash U CI-0014 US
; CURRENT APPLICATION NUMBER: US/09/104,070
; CURRENT FILING DATE: 1998-06-24
; EARLIER APPLICATION NUMBER: US 60/083,612
; EARLIER FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
;
; SEQ ID NO 1
; LENGTH: 5000
;
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (782)...(1252)
; NAME/KEY: CDS
; LOCATION: (1354)...(1440)
; NAME/KEY: CDS
; LOCATION: (1549)...(1895)
; NAME/KEY: CDS
; LOCATION: (1976)...(2165)
; NAME/KEY: CDS
; LOCATION: (2251)...(2437)
; NAME/KEY: CDS
; LOCATION: (2559)...(2629)
; NAME/KEY: CDS
; LOCATION: (2703)...(2892)
; NAME/KEY: CDS
; LOCATION: (2975)...(3070)
; NAME/KEY: CDS
; LOCATION: (3148)...(3242)
; NAME/KEY: CDS
; LOCATION: (3317)...(3436)
; NAME/KEY: CDS
; LOCATION: (3540)...(3659)
; NAME/KEY: CDS
; LOCATION: (3745)...(3843)
; NAME/KEY: CDS
; LOCATION: (3934)...(4038)
; NAME/KEY: CDS
; LOCATION: (4130)...(4354)
; NAME/KEY: CDS
; LOCATION: (4826)...(4755)
; OTHER INFORMATION: t-RNA-glu coding region
;
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (785)...(786)
; OTHER INFORMATION: some rearrangement, deletion of G at 785 and
; OTHER INFORMATION: insertion of 83 bp between 785 and 786
;
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (3243)...(3243)
; OTHER INFORMATION: ddm1-2 base pair substitution of G to A
;
; US-09-104-070-1
;
Query Match 64.0%; Score 16; DB 3; Length 5000;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 3 tgcgtgacgcgcgtt 18
|||||
Db 1132 tgcgtgacgcgcgtt 1147

```

```

RESULT 5
US-09-104-070-1
; Sequence 1, Application US/09104070
; Patent No. 6153741

```

```

RESULT 6
US-08-998-416-997/c
; Sequence 997, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:

```

APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Reibischung, Corline
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 623926artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CCG1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 997:
SEQUENCE CHARACTERISTICS:
LENGTH: 708 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1612RP
US-08-998-416-997

Query Match 60.0%; Score 15; DB 4; Length 708;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcctgagctcgtctgt 17
|||||

Db 470 TGCCTGAGCTCCTGTCTGT 456

RESULT 7
US-08-820-170A-8/C
Sequence 8, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.

COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
US-08-820-170A-8

Query Match 56.0%; Score 14; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcctcgtcgtctgtt 19
|||||

Db 350 TGCCTCCTCCTGTCTT 337

RESULT 8
US-09-055-699-8/C
Sequence 8, Application US/09055699
Patent No. 6005088
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,170
FILING DATE:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
US-09-055-699-8

Query Match 56.0%; Score 14; DB 3; Length 384;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgttt 19
|||||
DB 350 TGGATCTGCTGTTT 337

RESULT 9
US-09-273-565-8/c
Sequence 8, Application US/09273565A
Patent No. 6166190
GENERAL INFORMATION:
APPLICANT: FUJIMURA, TSUTOMU
APPLICANT: MATANABE, TAKESHI
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
FILE REFERENCE: Q-53599
CURRENT APPLICATION NUMBER: US/09/273,565A
EARLIER FILING DATE: 1999-03-22
EARLIER APPLICATION NUMBER: 09/055,699
EARLIER FILING DATE: 1998-04-07
EARLIER APPLICATION NUMBER: 08/820,170
EARLIER FILING DATE: 1997-03-19
EARLIER APPLICATION NUMBER: JP 63410/1996
EARLIER FILING DATE: 1996-03-19
EARLIER APPLICATION NUMBER: JP 69163/1997
EARLIER FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 384
TYPE: DNA
ORGANISM: Homo sapiens
US-09-273-565-8

Query Match 56.0%; Score 14; DB 4; Length 384;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgttt 19
|||||
DB 350 TGGATCTGCTGTTT 337

RESULT 10
PCT-US94-14106-54
Sequence 54, Application PC/TUS9414106
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Process for Generating Specific Antibodies
NUMBER OF SEQUENCES: 61
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/14106
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 base pairs
TYPE: nucleic acid

STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 67..735
PCT-US94-14106-54

Query Match 56.0%; Score 14; DB 5; Length 735;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcctgctgctg 16
|||||
DB 465 TGTGATCTGCTG 478

RESULT 11
US-09-232-191-14/c
Sequence 14, Application US/09232191
Patent No. 6284487
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
TITLE OF INVENTION: Fatty Acid Transport Proteins
FILE REFERENCE: WH197-21P3ME
CURRENT APPLICATION NUMBER: US/09/232,191
EARLIER FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 753
TYPE: DNA
ORGANISM: Homo sapiens
US-09-232-191-14

Query Match 56.0%; Score 14; DB 4; Length 753;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctgctgttgaag 23
|||||
DB 436 TCTGCTGTTGAG 423

RESULT 12
US-09-232-200-14/c
Sequence 14, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WH197-21P3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
EARLIER FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941

```
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-200-14
```

```
Query Match          56.0%; Score 14; DB 4; Length 753;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 tctgcgttgaag 23
      |||
Db 436 TCTGCTGTTGAAG 423
```

```
RESULT 13
US-09-232-197-14/c
; Sequence 14, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-197-14
```

```
Query Match          56.0%; Score 14; DB 4; Length 753;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 tctgcgttgaag 23
      |||
Db 436 TCTGCTGTTGAAG 423
```

```
RESULT 14
US-09-232-200-50/c
; Sequence 50, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
```

```
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 50
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-200-50
```

```
Query Match          56.0%; Score 14; DB 4; Length 1173;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 tctgcgttgaag 23
      |||
Db 853 TCTGCTGTTGAAG 840
```

```
RESULT 15
US-09-232-197-50/c
; Sequence 50, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 50
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-197-50
```

```
Query Match          56.0%; Score 14; DB 4; Length 1173;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 tctgcgttgaag 23
      |||
Db 853 TCTGCTGTTGAAG 840
```

```
RESULT 16
US-08-820-170A-9/c
; Sequence 9, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
```

STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
LIBRARY: Human fetal brain cDNA library
CLONE: GEN-025f07
FEATURE:
NAME/KEY: CDS
LOCATION: 95..478
US-08-820-170A-9

Query Match 56.0%; Score 14; DB 2; Length 1493;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 tggatctgctgttt 19
|||||
Db 444 TGGATCTGCTGTTT 431

RESULT 17
US-09-055-699-9/c
Sequence 9, Application US/09055699
Patent No. 6005088
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIMURA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/820,170
FILING DATE:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
LIBRARY: Human fetal brain cDNA library
CLONE: GEN-025f07
FEATURE:
NAME/KEY: CDS
LOCATION: 95..478
US-09-055-699-9

Query Match 56.0%; Score 14; DB 3; Length 1493;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 tggatctgctgttt 19
|||||
Db 444 TGGATCTGCTGTTT 431

RESULT 18
US-09-273-565-9/c
Sequence 9, Application US/09273565A
Patent No. 6166190
GENERAL INFORMATION:
APPLICANT: FUJIMURA, TSUTOMU
APPLICANT: WATANABE, TAKESHI
APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBQUITIN-CONJUGATING ENZYME
FILE REFERENCE: 0-53599
CURRENT APPLICATION NUMBER: US/09/273,565A
CURRENT FILING DATE: 1999-03-22
EARLIER APPLICATION NUMBER: 09/055,699
EARLIER FILING DATE: 1998-04-07
EARLIER APPLICATION NUMBER: 08/820,170
EARLIER FILING DATE: 1997-03-19
EARLIER APPLICATION NUMBER: JP 63410/1996
EARLIER FILING DATE: 1996-03-19
EARLIER APPLICATION NUMBER: JP 69163/1997
EARLIER FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 9
LENGTH: 1493
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (95)..(478)
US-09-273-565-9

Query Match 56.0%; Score 14; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 tggatctgctgttt 19
|||||
Db 444 TGGATCTGCTGTTT 431

RESULT 19
US-08-447-422-15/c
Sequence 15, Application US/08447422
Patent No. 5685579
GENERAL INFORMATION:
APPLICANT: SHAKI, Ezekiel Y.
APPLICANT: ROTHSTEIN, Aser
APPLICANT: RAMJESINGH, Mohabir
TITLE OF INVENTION: Use of Antibody/Antigen Interactions to
TITLE OF INVENTION: Protect or Modulate Biological Activity
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,422
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,410
FILING DATE: 22-JUN-1993
APPLICATION NUMBER: US 07/938,505
FILING DATE: 31-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/205,748
FILING DATE: 21-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17923/102 HYLI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ. ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-447-422-15

Query Match 56.0%; Score 14; DB 1; Length 1848;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tctgtgattctgctg 16
|||||

Db 1829 TCGCTGATCTGCTG 1816

RESULT 20
US-09-232-200-68/c
Sequence 68, Application US/09232200A
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS

FILE REFERENCE: WHI97-21P3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ. ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ. ID NO 68
LENGTH: 1998
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-200-68

Query Match 56.0%; Score 14; DB 4; Length 1998;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctgtgtttgaag 23
|||||

Db 1690 TCTGCTGTTGAG 1677

RESULT 21
US-09-232-197-68/c
Sequence 68, Application US/09232197A
Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ. ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ. ID NO 68
LENGTH: 1998
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-197-68

Query Match 56.0%; Score 14; DB 4; Length 1998;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctgtgtttgaag 23
|||||

Db 1690 TCTGCTGTTGAG 1677

RESULT 22
US-09-232-191-6/c
Sequence 6, Application US/09232191
Patent No. 6284487
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.

```

; TITLE OF INVENTION: Fatty Acid Transport Proteins
; FILE REFERENCE: WHI97-21p3ME
; CURRENT APPLICATION NUMBER: US/09/232,191
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2087
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-191-6
```

```

Query Match          56.0%; Score 14; DB 4; Length 2087;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 tctgctgttgaag 23
|||||
Db 1779 TCTGCTGTTGAAG 1766
```

```

RESULT 23
US-09-232-200-6/c
; Sequence 6, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2087
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-200-6
```

```

Query Match          56.0%; Score 14; DB 4; Length 2087;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 tctgctgttgaag 23
|||||
Db 1779 TCTGCTGTTGAAG 1766
```

```

RESULT 24
US-09-232-197-6/c
; Sequence 6, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
```

```

; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2087
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-197-6
```

```

Query Match          56.0%; Score 14; DB 4; Length 2087;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 tctgctgttgaag 23
|||||
Db 1779 TCTGCTGTTGAAG 1766
```

```

RESULT 25
US-09-232-200-101/c
; Sequence 101, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 2166
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)...(2124)
US-09-232-200-101
```

```

Query Match          56.0%; Score 14; DB 4; Length 2166;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 tctgctgttgaag 23
|||||
Db 1987 TCTGCTGTTGAAG 1974
```

```
RESULT 26
```

US-09-232-197-101/c
Sequence 101, Application US/09232197A
Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimco, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WH197-21P3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 101
LENGTH: 2166
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (19)...(2124)
US-09-232-197-101

Query Match 56.0%; Score 14; DB 4; Length 2166;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 tctgctgttgaag 23
|||||||
Db 1987 TCTGCTGTTTGAAG 1974

RESULT 27
US-08-239-276-6/c
Sequence 6, Application US/08239276
Patent No. 5840836
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/239,276
FILING DATE: 05-MAY-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703

FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3243 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-239-276-6

Query Match 56.0%; Score 14; DB 2; Length 3243;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 tctggtatctgctg 16
|||||||
Db 413 TCTGATCTCTGCTG 400

RESULT 28
US-08-468-579B-6/c
Sequence 6, Application US/08468579B
Patent No. 5981700
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,579B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.5-KGB

TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3243 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-468-579B-6

Query Match
Best Local Similarity 100.0%; Score 14; DB 2; Length 3243;
Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgcgtgacgcgcgc 16
|||||
Db 413 TGCTGATCTGCTG 400

RESULT 29
US-08-468-577B-6/c
Sequence 6, Application US/08468577B
Patent No. 6001804
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,577B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3243 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
US-08-468-577B-6

Query Match
Best Local Similarity 100.0%; Score 14; DB 3; Length 3243;
Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgcgtgacgcgcgc 16
|||||
Db 413 TGCTGATCTGCTG 400

RESULT 30
US-08-072-574-11/c
Sequence 11, Application US/08072574
Patent No. 5521297
GENERAL INFORMATION:
APPLICANT: Daggett, Lortie
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
TITLE OF INVENTION: HUMAN MATABOTROPIC GLUTAMATE RECEPTORS,
NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 00719
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/072,574
FILING DATE: 19930604
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen F.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9383
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 3282 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..3003
OTHER INFORMATION: /product= "HUMAN MGUR5C"
OTHER INFORMATION: /note= "Variant of MGUR5A with truncated 3' end."
US-08-072-574-11

Query Match
Best Local Similarity 100.0%; Score 14; DB 1; Length 3282;
Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgcgtgacgcgcgc 16
|||||
Db 2950 TGCTGATCTGCTG 2937

RESULT 31

US-08-486-270-11/C
; Sequence 11, Application US/08486270
; Patent No. 5807689
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,270
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: FP41 9772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-9392
; TELEFAX: 619-546-4737
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3003
; OTHER INFORMATION: /product="HUMAN MGLUR5C"
; OTHER INFORMATION: /note="Variant of MGLUR5A with truncated 3' end."
US-08-486-270-11

Query Match 56.0%; Score 14; DB 1; Length 3282;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgacgcgcg 16
|||||

Db 2950 TGCCTGATCTGCTG 2937

RESULT 32
US-08-367-264-11/C
; Sequence 11, Application US/08367264
; Patent No. 6001581
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.

APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/367,264
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: FP41 9772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3003
; OTHER INFORMATION: /product="HUMAN MGLUR5C"
; OTHER INFORMATION: /note="Variant of MGLUR5A with truncated 3' end."
US-08-367-264-11

Query Match 56.0%; Score 14; DB 3; Length 3282;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgacgcgcg 16
|||||

Db 2950 TGCCTGATCTGCTG 2937

RESULT 33
US-08-072-574-7/C
; Sequence 7, Application US/08072574
; Patent No. 5521297
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 00719

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/072,574
? FILING DATE: 19930604
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Reiter, Stephen E.
? REGISTRATION NUMBER: 31,192
? REFERENCE/DOCKET NUMBER: P41 9383
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 213-622-7700
? TELEFAX: 213-489-4210
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4085 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 370..3912
? OTHER INFORMATION: /product= "HUMAN MGLUR5A"
?
US-08-072-574-7

Query Match          56.0%; Score 14; DB 1; Length 4085;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgcgtgacgtcgtg 16
Db 2950 TGCTGATCTGCTG 2937

RESULT 34
US-08-486-270-7/c
? Sequence 7, Application US/08486270
? Patent No. 5807689
? GENERAL INFORMATION:
? APPLICANT: Daggett, Lorrie
? APPLICANT: Ellis, Steven B.
? APPLICANT: Liaw, Chen
? APPLICANT: Pontsler, Aaron
? APPLICANT: Johnson, Edwin C.
? APPLICANT: Hess, Stephen D.
? TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
? TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
? NUMBER OF SEQUENCES: 13
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
? STREET: 444 South Flower Street, Suite 2000
? CITY: Los Angeles
? STATE: CA
? COUNTRY: USA
? ZIP: 90071
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/486,270
? FILING DATE: 02-JUN-1994
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/072,574
? FILING DATE: 04-JUN-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Reiter, Stephen E.
? REGISTRATION NUMBER: 31,192
? REFERENCE/DOCKET NUMBER: P41 9772
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 619-546-4737
? TELEFAX: 619-546-9392
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4085 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 370..3912
? OTHER INFORMATION: /product= "HUMAN MGLUR5A"
?
US-08-486-270-7
```

```

? NAME: Reiter, Stephen E.
? REGISTRATION NUMBER: 31,192
? REFERENCE/DOCKET NUMBER: P41 9772
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 619-546-4737
? TELEFAX: 619-546-9392
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4085 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 370..3912
? OTHER INFORMATION: /product= "HUMAN MGLUR5A"
?
US-08-486-270-7

Query Match          56.0%; Score 14; DB 1; Length 4085;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgcgtgacgtcgtg 16
Db 2950 TGCTGATCTGCTG 2937

RESULT 35
US-08-367-264-7/c
? Sequence 7, Application US/08367264
? Patent No. 6001581
? GENERAL INFORMATION:
? APPLICANT: Daggett, Lorrie
? APPLICANT: Ellis, Steven B.
? APPLICANT: Liaw, Chen
? APPLICANT: Pontsler, Aaron
? APPLICANT: Johnson, Edwin C.
? APPLICANT: Hess, Stephen D.
? TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
? TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
? NUMBER OF SEQUENCES: 13
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
? STREET: 444 South Flower Street, Suite 2000
? CITY: Los Angeles
? STATE: CA
? COUNTRY: USA
? ZIP: 90071
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/367,264
? FILING DATE: 02-JUN-1994
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/072,574
? FILING DATE: 04-JUN-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Reiter, Stephen E.
? REGISTRATION NUMBER: 31,192
? REFERENCE/DOCKET NUMBER: P41 9772
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 619-546-4737
? TELEFAX: 619-546-9392
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4085 base pairs
? TYPE: nucleic acid
```

STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..3912
OTHER INFORMATION: /product= "HUMAN MGLUR5A"
US-08-367-264-7

Query Match 56.0%; Score 14; DB 3; Length 4085;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgctgacatctgctg 16
|||||
Db 2950 TGCTGATCTGCTG 2937

RESULT 36

US-08-072-574-9/c
Sequence 9, Application US/08072574
Patent No. 5521297

GENERAL INFORMATION:
APPLICANT: Daggett, Lorrle
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 00719

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/072,574
FILING DATE: 19930604
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9383
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /product= "HUMAN MGLUR5H"
OTHER INFORMATION: /note= "Variant of MGLUR5A with 96 base pair
OTHER INFORMATION: Insertion between nucleotides 2999 and 3000."
US-08-072-574-9

Query Match 56.0%; Score 14; DB 1; Length 4181;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgctgacatctgctg 16
|||||
Db 2950 TGCTGATCTGCTG 2937

RESULT 37

US-08-486-270-9/c
Sequence 9, Application US/08486270
Patent No. 5807689

GENERAL INFORMATION:
APPLICANT: Daggett, Lorrle
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
APPLICANT: Johnson, Edwin C.
APPLICANT: Hess, Stephen D.
TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,270
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /product= "HUMAN MGLUR5B"
OTHER INFORMATION: /note= "Variant of MGLUR5A with 96 base pair
OTHER INFORMATION: Insertion between nucleotides 2998 and 2999."
US-08-486-270-9

Query Match 56.0%; Score 14; DB 1; Length 4181;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgctgacatctgctg 16
|||||
Db 2950 TGCTGATCTGCTG 2937

RESULT 38

US-08-367-264-9/c

Sequence 9, Application US/08367264
Patent No. 6001581
GENERAL INFORMATION:
APPLICANT: Daggett, Lorrie
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Fontisler, Aaron
APPLICANT: Johnson, Edwin C.
APPLICANT: Hess, Stephen D.
TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/367,264
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /product= "HUMAN MGLUR5B"
OTHER INFORMATION: /note= "Variant of MGLUR5A with 96 base pair
OTHER INFORMATION: Insertion between nucleotides 2998 and 2999."
US-08-367-264-9

Query Match 56.0%; Score 14; DB 3; Length 4181;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgactcgtc 16
|||||

Db 2950 TCGTGACTGCTG 2937

RESULT 39
US-08-660-148-1/c
Sequence 1, Application US/08660148
Patent No. 6211353
GENERAL INFORMATION:
APPLICANT: Burnett, J. P.
APPLICANT: Mayne, Nancy G.
APPLICANT: Sharp, Robert L.
APPLICANT: Snyder, Yvonne M.
TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND

TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4207 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 460..3999
US-08-660-148-1

Query Match 56.0%; Score 14; DB 4; Length 4207;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgactcgtc 16
|||||

Db 3040 TCGTGACTGCTG 3027

RESULT 40
US-08-660-148-3/c
Sequence 3, Application US/08660148
Patent No. 6211353
GENERAL INFORMATION:
APPLICANT: Burnett, J. P.
APPLICANT: Mayne, Nancy G.
APPLICANT: Sharp, Robert L.
APPLICANT: Snyder, Yvonne M.
TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4207 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
US-08-660-148-3

Query Match 56.0%; Score 14; DB 4; Length 4207;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgactgctg 16
|||||
Db 3040 TGCTGATCTGCTG 3027

RESULT 41
US-08-660-148-4/c
Sequence 4, Application US/08660148
Patent No. 6211353
GENERAL INFORMATION:
APPLICANT: Burnett, J. P.
APPLICANT: Mayne, Nancy G.
APPLICANT: Sharp, Robert L.
APPLICANT: Snyder, Yvonne M.
TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 4303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 460..4095
US-08-660-148-4

Query Match 56.0%; Score 14; DB 4; Length 4303;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgactgctg 16
|||||
Db 3040 TGCTGATCTGCTG 3027

RESULT 42
US-08-660-148-6/c
Sequence 6, Application US/08660148
Patent No. 6211353
GENERAL INFORMATION:

APPLICANT: Burnett, J. P.
APPLICANT: Mayne, Nancy G.
APPLICANT: Sharp, Robert L.
APPLICANT: Snyder, Yvonne M.
TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
US-08-660-148-6

Query Match 56.0%; Score 14; DB 4; Length 4303;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tctgagctctgctg 16
|||||
Db 3040 TCGGATCTGCTG 3027

RESULT 43

US-08-465-976A-1/C
; Sequence 1, Application US/08465976A
; Patent No. 5869632
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; APPLICANT: RUBEN, STEVEN M
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN GILLILLIAN, CECCHI
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: US
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,976A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY F
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-444
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 994-1700
; TELEFAX: (201) 994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 212..2863
US-08-465-976A-1

Query Match 56.0%; Score 14; DB 2; Length 4566;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgagatctgctgt 18
|||||
Db 4182 CTGGATCTGCTGT 4169

RESULT 44

US-08-982-412-1/C
; Sequence 1, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; APPLICANT: RUBEN, STEVEN M
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
;; STREET: 9410 KEY WEST AVENUE
;; CITY: ROCKVILLE,
;; STATE: MD
;; COUNTRY: US
;; ZIP: 20850

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/982,412
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BROOKES, ANDERS A
;; REGISTRATION NUMBER: 36,373
;; REFERENCE/DOCKET NUMBER: PF181PCT2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (301) 309-8504
;; TELEFAX: (301) 309-8439
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4566 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 212..2863
US-08-982-412-1

Query Match 56.0%; Score 14; DB 2; Length 4566;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgagatctgctgt 18
|||||
Db 4182 CTGGATCTGCTGT 4169

RESULT 45

US-08-169-927-1
; Sequence 1, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dodson, Michael E.
; APPLICANT: Ching, Wei Mei
; APPLICANT: Dasch, Gregory A
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the detection of both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:

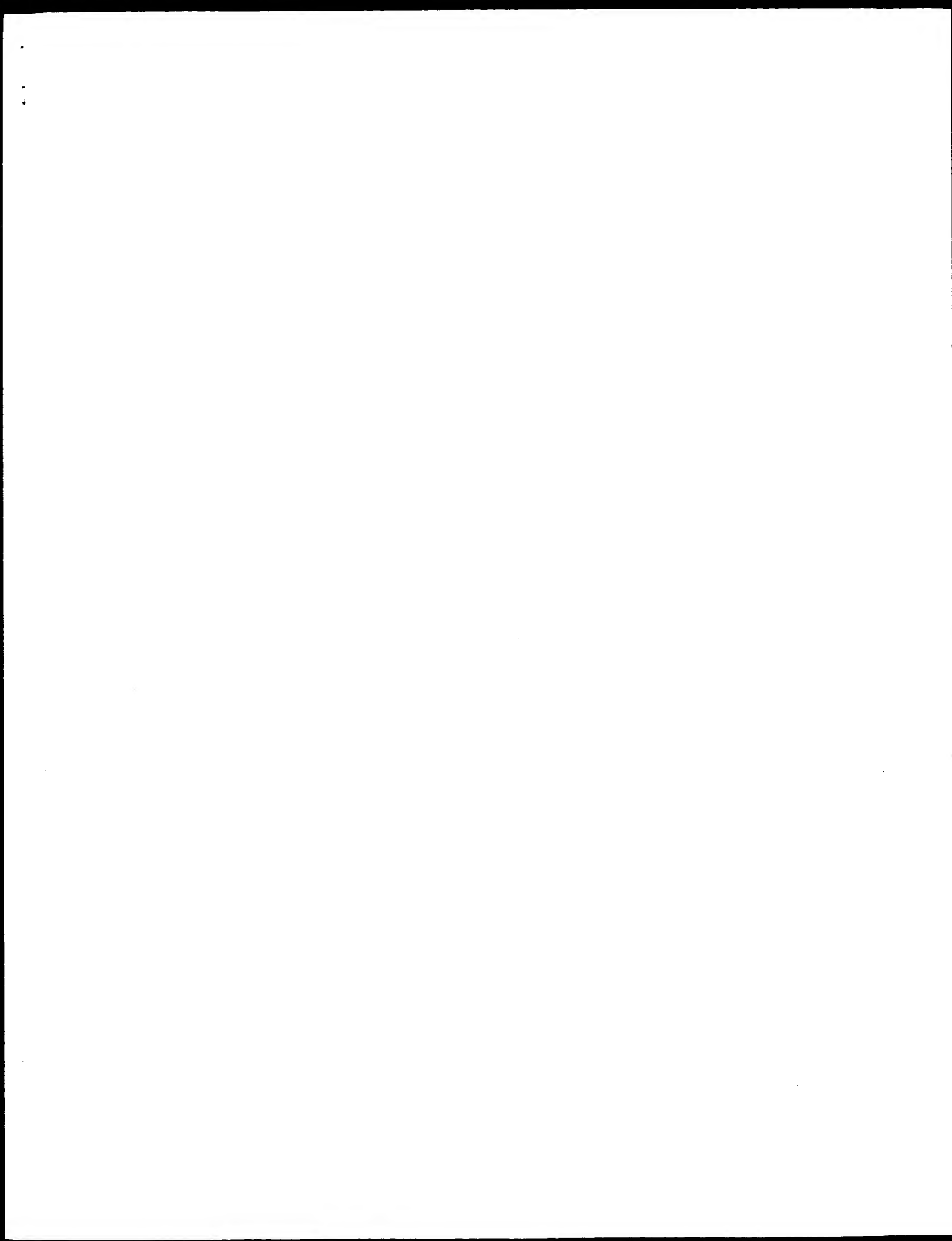
```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5319 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Rickettsia prowazekii
STRAIN: Breinl
FEATURE:
NAME/KEY: -35_signal
LOCATION: 340..345
FEATURE:
NAME/KEY: -10_signal
LOCATION: 363..368
FEATURE:
NAME/KEY: CDS
LOCATION: 391..5226
FEATURE:
NAME/KEY: RBS
LOCATION: 379..386
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5270..5306
PUBLICATION INFORMATION:
AUTHORS: Carl, M.
AUTHORS: Dobson, M. E.
AUTHORS: Ching, W. M.
AUTHORS: Dasch, G. A.
TITLE: Characterization of the gene encoding the
TITLE: protective S-layer protein of Rickettsia
TITLE: prowazekii; presence of a truncated identical
TITLE: homolog in rickettsia typhi
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
DATE: 1990
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
US-08-169-927-1

Query Match 56.0%; Score 14; DB 1; Length 5319;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgtgagatcgc 14
|||||
Db 4321 gttgtgagatcgc 4334

```

Search completed: December 26, 2001, 12:52:35
 Job time: 6787 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:09:56 : Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196f-3

Perfect score: 25

Sequence: 1 gtgtcgtgactcgtcgttgaagcg 25

Scoring table: OLIGO_NUC

Searched: 11351937 seqs, 5372889281 residues

Word size : 9

Total number of hits satisfying chosen parameters: 862398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : EST:*

1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_iny:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Match	Query Length	ID	Description
1	19	76.0	888 13	CNS04KH2
2	18	72.0	403 10	BE095496
3	18	72.0	487 11	BF428015
4	17	68.0	546 13	AO952526
5	17	68.0	547 10	AV718673
6	16	64.0	237 10	BE023256
7	16	64.0	274 10	BE023256
8	16	64.0	274 11	BF414722
9	16	64.0	287 10	BF455974
10	16	64.0	287 11	BF059872
11	16	64.0	287 11	BF917535
12	16	64.0	293 13	AZ214122

13	16	64.0	301 10	AM306120
14	16	64.0	328 10	BE659333
15	16	64.0	342 10	AM423279
16	16	64.0	360 11	CE1070
17	16	64.0	365 10	BE662000
18	16	64.0	412 11	BE1054795
19	16	64.0	422 10	AT1736967
20	16	64.0	434 10	AM704839
21	16	64.0	436 10	AT1293622
22	16	64.0	441 10	BE190258
23	16	64.0	470 10	AL1373871
24	16	64.0	474 10	AM704596
25	16	64.0	481 10	AL1379973
26	16	64.0	497 11	BF461669
27	16	64.0	502 10	BE515828
28	16	64.0	503 10	AL383090
29	16	64.0	513 11	BF148981
30	16	64.0	517 10	AL1373872
31	16	64.0	517 13	AO692054
32	16	64.0	521 11	BF419551
33	16	64.0	522 10	AT854720
34	16	64.0	525 10	AL1379974
35	16	64.0	531 10	AM935303
36	16	64.0	555 13	AO889100
37	16	64.0	570 11	BI229421
38	16	64.0	574 10	AM573807
39	16	64.0	575 11	BE999070
40	16	64.0	580 11	BE525150
41	16	64.0	591 10	AM734894
42	16	64.0	601 13	AZ267971
43	16	64.0	602 11	BI229406
44	16	64.0	609 10	AM267738
45	16	64.0	616 10	AM980891
46	16	64.0	618 11	BE585528
47	16	64.0	621 10	AM422111
48	16	64.0	627 13	BE593142
49	16	64.0	632 10	AZ449344
50	16	64.0	634 11	BE634485
51	16	64.0	640 10	AT108765
52	16	64.0	643 13	AZ015305
53	16	64.0	650 13	1A69C05P
54	16	64.0	755 10	BE034735
55	16	64.0	831 11	BE584066
56	16	64.0	834 13	CNS01V13
57	16	64.0	1104 13	CNS02ARM
58	16	64.0	1177 11	BI389308
59	16	64.0	177 13	FR0044714
60	16	64.0	190 11	BI024308
61	16	64.0	196 10	BE398642
62	16	64.0	227 10	BE012248
63	16	64.0	243 10	AT547503
64	16	64.0	277 11	BE980507
65	16	64.0	278 11	AT1021028
66	16	64.0	295 11	BE440174
67	16	64.0	298 10	BE337949
68	16	64.0	300 11	C53496
69	16	64.0	300 11	C51717
70	16	64.0	300 11	C52955
71	16	64.0	300 11	C59449
72	16	64.0	301 11	BE990261
73	16	64.0	303 11	BE990318
74	16	64.0	316 13	FR0039786
75	16	64.0	342 10	AV2009024
76	16	64.0	348 13	AZ228552
77	16	64.0	356 11	BE375907
78	16	64.0	358 11	BE355075
79	16	64.0	360 10	AT486110
80	16	64.0	372 13	AO521765
81	16	64.0	398 11	BI389121
82	16	64.0	403 13	AM677232
83	16	64.0	406 13	AO143500
84	16	64.0	424 13	AO494880
85	16	64.0	427 13	BH077581

```

86 15 60.0 435 13 AO114851 C1T-HSP-2
87 15 60.0 437 10 AW432846 sb1905.Y
88 15 60.0 442 10 AA397373 nc64c07.s
89 15 60.0 443 11 BF415104 UI-R-BJ2-
90 15 60.0 445 10 BE580608 kp79911.Y
91 15 60.0 456 11 BE555582 UI-R-E0-C
92 15 60.0 458 11 BG379251 UI-R-BT1-
93 15 60.0 479 13 AZ965014 A2965014 2M0234K24
94 15 60.0 489 13 A1958363 f603904.Y
95 15 60.0 495 13 CNS04S8M A1304879 Tetraodon
96 15 60.0 495 13 AO192174 HS.3227.B
97 15 60.0 501 10 A1984034 A1984034 M552f02.X
98 15 60.0 507 10 AM660235 98156 MAR
99 15 60.0 507 10 BE114310 UI-R-BJ1-
C 100 15 60.0 509 10 AW915326 ESW346630

```

ALIGNMENTS

```

RESULT 1
LOCUS CNS04KH2 888 bp DNA GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence.T7 end of clone
116p03 of library G from Tetraodon nigroviridis, genomic survey
sequence.

```

```

ACCESSION AL294815.1 GI:8033395
VERSION GSS: genome survey sequence.
KEYWORDS Tetraodon nigroviridis.
SOURCE Tetraodon nigroviridis.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
Tetraodontidae; Tetraodon.

```

```

REFERENCE 1 (bases 1 to 888)
AUTHORS Roest-Crollius,H., Jallou,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

```

```

TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 888)
AUTHORS Roest-Crollius,H., Jallou,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

```

```

TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

```

```

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 888)
AUTHORS Direct Submission

```

```

COMMENT Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

```

```

FEATURES
source 1..888
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="G"
/clone="116D03"
/note="Genoscope sequence ID : COB0116CR02LPI-end : 77"

```

```

BASE COUNT 129 a 276 c 276 g 187 t 20 others
ORIGIN

```

```

Query Match 76.0%; Score 19; DB 13; Length 888;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 3 tgcgtgactgctgttga 21

```

```

Db 331 TGCCTGATCTGCTGTTGA 349

```

```

RESULT 2
LOCUS BE095496/c 403 bp mRNA EST 12-JUN-2000
DEFINITION UI-R-BU0-apa-c-11-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone
UI-R-BU0-apa-c-11-0-UI.3', mRNA sequence.

```

```

ACCESSION BE095496
VERSION BE095496
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE 1 (bases 1 to 403)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477

```

```

COMMENT Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscores@blue.weeg.uiowa.edu
Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 forward
POLYA=No.

```

```

FEATURES
source 1..403
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BU0-apa-c-11-0-UI"
/clone_lib="UI-R-BU0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The UI-R-BU0
library is a subtracted library derived from a mixture of
eye and ganglia tissues. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratseq.eng.uiowa.edu. The subtraction has
been previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG-Seq=None found"

```

```

BASE COUNT 85 a 147 c 141 g 30 t
ORIGIN

```

```

Query Match 72.0%; Score 18; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 gtcgtgactgctgtt 18
Db 245 GTTCCTGATCTGCTGTT 228

```

```

RESULT 3
LOCUS BF428015 487 bp mRNA EST 30-MAR-2001
DEFINITION daa04g08.x1 NICHD XCC OVI Xenopus laevis cDNA clone IMAGE:4055247
3' similar to TR:093478 093478 KINESTIN LIKE PROTEIN 3.; mRNA
sequence.

```

```


```

ACCESSION BF428015
 VERSION BF428015.1 GI:11439676
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM *Xenopus laevis*
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 Xenopodinae; *Xenopus*.

REFERENCE 1 (bases 1 to 487)
 Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person, B., Gibbons, M., Harvey, N., Rilter, E., Jackson, Y., McCann, R., Westerson, R. and Wilson, R.
 Washu *Xenopus* EST project, 1999
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-f@mail.nih.gov
 Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: *Xenopus* clones from this library are available through the I.M.A.G.E. Consortium/BLNL at: infoimage.llnl.gov
 Seq primer: -400P from Gibco
 High quality sequence stop: 50.
 Location/Qualifiers

FEATURES
 source
 1..487
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone_image="405247"
 /clone_lib="NICH XGC OVI"
 /sex="female"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pCMV-SPOK16; Site: 1; NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.0 kb. Constructed by Life Technologies."
 Technology.

BASE COUNT 138 a 101 c 102 g 146 t
 ORIGIN

Query Match 72.0%; Score 18; DB 11; Length 487;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgacatcgtctgt 18
 ||||||||||||||||
 Db 249 GTTGCTGCATCTGCTGT 266

RESULT 4
 A0952526 546 bp DNA GSS 27-JAN-2000
 LOCUS A0952526
 DEFINITION Sheared DNA-48D19.1F Sheared DNA Trypanosoma brucei genomic clone
 A0952526
 ACCESSION A0952526
 VERSION A0952526.1 GI:6775791
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma

REFERENCE 1 (bases 1 to 546)
 El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J., Fraser, C. and Adams, M.
 Determination of clone end sequences from *Trypanosoma brucei* GUTat 10.1 sheared DNA library
 Unpublished (1999)
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200
 Fax: 301 838 0208
 Email: neisayed@tigr.org
 Clones are derived from the *Trypanosoma brucei* GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/tdb/mdb/tbdb/>.
 Seq primer: M13-Forward
 Class: shotgun.

FEATURES
 source
 1..546
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-48D19"
 /clone_lib="Sheared DNA"
 /note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of *Trypanosoma brucei* (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 192 a 161 c 127 g 66 t
 ORIGIN

Query Match 68.0%; Score 17; DB 13; Length 546;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgacatcgtctgt 17
 ||||||||||||||||
 Db 192 GTTGCTGCATCTGCTGT 176

RESULT 5
 AV718673 547 bp mRNA EST 16-OCT-2000
 LOCUS AV718673
 DEFINITION AV718673 GLC Homo sapiens CDNA clone GLCDBA12.5', mRNA sequence.
 AV718673
 ACCESSION AV718673.1 GI:10815825
 VERSION AV718673.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 547)
 Qian, B., Wu, T., Peng, Q., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H., Xu, X., Li, N., Huang, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gu, Y., Chen, Z., and Han, Z.
 Homo sapiens CDNA GLC clones
 Unpublished (2000)
 Contact: Zengqiang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
 Tel: 86-21-50801919 (ex.45)
 Fax: 86-21-50801922
 Email: hanzq@ncgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 1..547
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GLCDBA12"
 /clone_lib="GLC"
 /tissue_type="corresponding non cancerous liver tissue"

```

/dev_stage="Adult"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      183 a      101 c      116 g      147 t
ORIGIN

Query Match      68.0%; Score 17; DB 10; Length 547;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 gttcgtgactgctgt 17
        |||
Db      473 GTTGTGATCTGCTGT 457

RESULT 6
LOCUS   AM101291      237 bp      mRNA      EST      06-DEC-1999
DEFINITION   sd77b02.y1 Gm-cl1009 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl1009-388 5' similar to TR:Q40380 Q40380 ARABINOGALACTAN-PROTEIN
PRECURSOR. ;, mRNA sequence.
ACCESSION   AM101291
VERSION     AM101291.1 GI:6071904
KEYWORDS    EST.
SOURCE      soybean.
ORGANISM    Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1 (bases 1 to 237)
AUTHORS    Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
TITLE      Public Soybean EST Project
JOURNAL    Unpublished (1999)
COMMENT    Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco.

FEATURES
Source
    Location/Qualifiers
        1..237
        /organism="Glycine max"
        /db_xref="taxon:3847"
        /clone="GENOME SYSTEMS CLONE ID: Gm-cl1009-388"
        /clone_1ib="Gm-cl1009"
        /lab_host="XLI10-GOLD"
        /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
        XhoI; The mRNA was isolated from entire roots of
        2-month-old 'Williams' plants that were greenhouse grown
        in 5-gallon pots. To suppress nodulation, black gold
        All-Purpose potting soil was supplemented with: 0.36g/L
        available phosphoric acid (P205), 20mg/L urea N, 0.16g/L
        S, 0.49mg/L B, 2.5mg/L Cu, 0.15g/L Fe, 13.53mg/L Mn,
        0.26mg/L Mo, 14mg/L Zn, 20mg/L Ca, and the following
        nutrients in a slow-release form (Osmocote): 0.155g/L
        ammonia N, 0.185g/L nitrate N, 0.35g/L available
        phosphoric acid, and 0.35g/L soluble potash. No nodules
        were visible on the roots at harvest. StratiGene's cDNA

```

```

Synthesis Kit (catalog #200401) was used to synthesize the
cDNA. First-strand synthesis was performed with 5-methyl
dCMP, hence the ligated cDNA is hemimethylated.
StratiGene's first-strand synthesis primer was used
(GAGAGAGAGAGAGAGAGAACTACTCTGAG(T)-18). After
second-strand synthesis is, the cDNA ends were 'polished'
with clone Pfu DNA polymerase, ligated to EcoRI adapters,
and phosphorylated. The XhoI site within the first-strand
synthesis primer was restricted by digestion with XhoI;
all XhoI sites in the cDNA would be protected by their
hemimethylated status. The cDNA constructs were
size-fractionated with a 400bp cutoff, using a SizeSep 400
Spin column from Pharmacia. The column eluent was then
ligated into StratiGene's pBluescript II XR predigested
vector (pBluescript II SK(+)) that had been digested with
EcoRI and XhoI, and phosphorylated. Both the white and
blue colonies appear to contain recombinant plasmids with
cDNA inserts. This library was constructed by Dr. Paul
Keim and Dr. Virginia Coryell."
BASE COUNT      40 a      60 c      52 g      85 t
ORIGIN

Query Match      64.0%; Score 16; DB 10; Length 237;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 gttcgtgactgctgt 16
        |||
Db      96 GTTGTGATCTGCTGT 111

RESULT 7
LOCUS   BE023256      247 bp      mRNA      EST      21-NOV-2000
DEFINITION   sm80a02.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl015-6339 5', mRNA sequence.
ACCESSION   BE023256
VERSION     BE023256.1 GI:8285697
KEYWORDS    EST.
SOURCE      soybean.
ORGANISM    Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1 (bases 1 to 247)
AUTHORS    Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
TITLE      Public Soybean EST Project
JOURNAL    Unpublished (1999)
COMMENT    Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Insert Length: 438 Std Error: 0.00
            Seq primer: -40RP from Gibco.

FEATURES
Source
    Location/Qualifiers
        1..247
        /organism="Glycine max"
        /db_xref="taxon:3847"

```



```

/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-6339"
/clone_lib="Gm-cl015"
/tissue_type="Mature flowers, field grown plants"
/lab_host="XL10-Gold"
/notes="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from mature flowers of field grown plants. The cDNA
library was prepared using the Stratagene pBluescript II
XR cDNA library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelting."

```

BASE COUNT 39 a 65 c 57 g 77 t 9 others

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 247;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gttgctgacatctgctg 16
 |||||
 Db 128 GTTGCTGATCTGCTG 143

RESULT 8
 LOCUS BF14722 274 bp mRNA EST 28-NOV-2000
 DEFINITION UI-R-BJ2-bow-c-11-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone
 UI-R-BJ2-bow-c-11-0-UI 3', mRNA sequence.
 ACCESSION BF14722
 VERSION BF14722.1 GI:11402711
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 274)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL 97044477
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Oligo-dT track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA library Preparation: M.B.
 Soares Lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA-No

FEATURES
 SOURCE
 1. 274
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BJ2-bow-c-11-0-UI"
 /clone_lib="UI-R-BJ2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2

library is a subtracted library derived from the following
 tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc,
 atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15
 dpc, AV canal at 15 dpc. For a detailed description of
 the library from which this clone was derived, please
 visit our web site at ratist.eng.uiowa.edu. The
 subtraction has been previously described in (Bonaldo,
 Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_SEQ=None found"

BASE COUNT 54 a 96 c 106 g 18 t

ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 274;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gttgctgacatctgctg 16
 |||||
 Db 246 GTTGCTGATCTGCTG 231

RESULT 9
 LOCUS BB455974 287 bp mRNA EST 21-JUL-2000
 DEFINITION BB455974 RIKEN full-length enriched, 12 days embryo spinal ganglion
 Mus musculus cDNA clone D130043H15 3', mRNA sequence.
 ACCESSION BB455974
 VERSION BB455974.1 GI:9351467
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 287)
 Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arahawa,T., Carninci,
 Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arahawa,T., Carninci,
 P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
 Hirozane,T., Horii,F., Ishii,Y., Ishikawa,J., Ishikawa,N., Itoh,M.,
 Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,T., Kikuchi,N.,
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
 Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,
 Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
 Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,T., Tomioka,N., Toya,
 T., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yamazaki,I.,
 Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,
 M., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Konno,H., et al.)
 Unpublished (2000)
 RIKEN Mouse ESTs (Konno,H., et al.)
 Contact: Yoshinori Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gs.c.riken.go.jp
 URL: <http://genome.gsc.riken.go.jp/>
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Thermostabilization and thermocycling of the thermolabile enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh,M., Kusunagi,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,T.,
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,
 Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 63 a 62 c 97 g 65 t

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 9 atctgctgtttgaac 24
Db 149 ATCTGCTGTTGAAC 164

Query Match 64.0%; Score 16; DB 11; Length 287;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtcgtgcatctgctg 16
Db 91 GTTGCTGCATCTGCTG 76

RESULT 13
LOCUS AM306120 301 bp mRNA EST 20-JAN-2000
DEFINITION se46b06.y1 Gm-c1017 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1017-1932 5', mRNA sequence.

ACCESSION AM306120
VERSION AM306120.1 GI:6718473
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE 1 (bases 1 to 301)
AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
R., Ritzer,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Wilson,R.
Public Soybean EST Project

TITLE Unpublished (1999)
JOURNAL Contact: Shoemaker R/Public Soybean EST Project
COMMENT Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from gibco.

FEATURES

source

1. 301
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1017-1932"
/clone_1ib="Gm-c1017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="X110-Gold"
/note="Vector: pBluescript II XR; Site: 1: EcoRI; Site: 2:
XhoI; This cDNA library was constructed from mRNA isolated
from vegetable buds of field grown plants. The cDNA
library was prepared using the Stratagene pBluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into X110-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelting."

BASE COUNT 73 a 48 c 60 g 119 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 301;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 63 a 62 c 97 g 65 t

Query Match 64.0%; Score 16; DB 11; Length 287;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtcgtgcatctgctg 16
Db 91 GTTGCTGCATCTGCTG 76

RESULT 12
LOCUS A2214122 293 bp DNA GSS 09-JUN-2000
DEFINITION Sheared DNA-67G1.TF Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-67G1, DNA sequence.
ACCESSION A2214122 GI:8431922
VERSION A2214122.1
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 293)
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,T., Fujii,C.,
Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other_GSS: Sheared DNA-67G1.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through Research Genetics, Alabama, USA. Sheared DNA
end sequences search page: <http://www.tigr.org/tdb/mbd/tbdb/>.
Seq primer: M13-Forward
Class: Shotgun.

FEATURES

source

1. 293
Location/Qualifiers
/organism="Trypanosoma brucei"
/strain="TREN927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-67G1"
/clone_1ib="Sheared DNA"
/note="Vector: pUC18; Site: 1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREN927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. in Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Borell, Oxford University
Press, 1999)."

BASE COUNT 38 a 73 c 78 g 104 t

ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 293;
Best Local Similarity 100.0%; Pred. No. 1e+02;

Query Match 64.0%; Score 16; DB 10; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gctgcgtgctctg 16
 |||
 DB 101 GTTGTGATCTGCTG 116

RESULT 16
 C61070/c 360 bp mRNA EST 22-SEP-1997
 LOCUS C61070 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
 DEFINITION clone yk224c7 5', mRNA sequence.
 ACCESSION C61070
 VERSION C61070.1 GI:2419775
 KEYWORDS EST.
 ORGANISM Caenorhabditis elegans.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea
 ; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 360)
 AUTHORS Kohara, Y., Motoshashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano
 M., Miyata, A. and Nishigaki, A.
 TITLE Expression map of the C.elegans genome
 JOURNAL Unpublished (1996)
 COMMENT Contact: Yuji Kohara
 Genome Biology Lab.
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.
 Location/Qualifiers
 1..360
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone_lib="Yuji Kohara unpublished cDNA"
 /sex="hermaphrodite, male"
 /tissue_type="whole animal"
 /dev_stage="varied"
 /dev_stage="varied"

BASE COUNT 101 a 71 c 76 g 111 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 360;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 auctgtgttgaagc 24
 |||
 DB 26 ATCTGCTGTTGAGC 11

RESULT 17
 BE662000/c 365 bp mRNA EST 08-JAN-2001
 LOCUS BE662000 Drosophila melanogaster adult testis library Drosophila
 DEFINITION bs02d05.y1 Drosophila melanogaster adult testis library Drosophila
 melanogaster cDNA clone bs02d05 5', mRNA sequence.
 ACCESSION BE662000
 VERSION BE662000.1 GI:9992572
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 365)
 AUTHORS Andrews, J., Bouffard, G.G., Chandle, C., Lu, J., Becker, K.G. and

TITLE Oliver, B.
 JOURNAL Gene discovery using computational and microarray analysis of
 MEDLINE transcription in the drosophila melanogaster testis
 COMMENT Genome Res. 10 (12), 2050-2043 (2000)
 20568492
 Contact: Brian Oliver
 Laboratory of Cellular and Developmental Biology
 NIDDK, National Institutes of Health
 6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
 Fax: (301) 496 5239
 Email: oliver@helix.nih.gov,
 http://www.niddk.nih.gov/Intram/people/poliver.htm
 Tissue isolation and library construction performed at the National
 Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
 http://www.niddk.nih.gov/Intram/people/poliver.htm). DNA sequencing
 and analyses performed by National Institutes of Health Intramural
 Sequencing Center (NISC; see http://www.nisc.nih.gov).
 Plate: 02 row: d column: 05
 Seq primer: M13RP1 reverse primer (ABI).
 Location/Qualifiers
 1..365
 /organism="Drosophila melanogaster"
 /strain="y[*] w[67c1]/y"
 /db_xref="taxon:7227"
 /clone_lib="bs02d05"
 /clone_lib="Drosophila melanogaster adult testis library"
 /sex="male"
 /dev_stage="1-5 day adult"
 /lab_host="SOLR (Stratagene)"
 /note="Organ: testis; Vector: pBluescript SK (Stratagene);
 Site: 1: EcoR I; Site: 2: Xho I; Testes dissected from 1-5
 day adult y[*] w[67c1]/y males raised at 25oc. RNA
 isolated using Trizol (Life Technologies) and a single
 round of Poly(A)+ selection using Oligotex (Qiagen). cDNA
 library constructed using Stratagene ZAP-cDNA synthesis
 kit. Oligo dt-primed, size fractionated -1-6 kb, and
 directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
 Following a single round of amplification pBluescript SK
 phagemids were mass excised. A distribution channel for
 clones is being sought, but not currently available.
 Requests for clones cannot be honored."

BASE COUNT 85 a 129 c 94 g 56 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 365;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gctgcgtgctctg 16
 |||
 DB 330 GTTGTGATCTGCTG 315

RESULT 18
 B1054795/c 412 bp mRNA EST 15-JUN-2001
 LOCUS B1054795 MR3-GN0349-020201-003-f12 GN0349 Homo sapiens cDNA, mRNA sequence.
 DEFINITION B1054795
 ACCESSION B1054795
 VERSION B1054795.1 GI:14462325
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 412)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalhal, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.J.
 Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL
MEDLINE
COMMENT

sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR3&cl2=MR3-GN0349-
020201-003-f12&ts=2001-02-02&ft=1)
Seq primer: puc 18 forward

High quality sequence stop: 412.

FEATURES

source

1.412
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0349"
/dev_stage="Adult"
/note="Organ: placenta.normal; Vector: puc18; Site_1: SmaI
/Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
Application No. 196/716 - Ludwig Institute for Cancer
Research) profiles into the puc 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 102 a 117 c 107 g 86 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 412;
Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgagctctgctg 16
|||||

Db 397 gttgctgagctctgctg 382

RESULT 19

LOCUS A1736967 422 bp mRNA EST 31-JUL-2001
DEFINITION SB35C02.Y1 Gm-c1013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1013-27 5' similar to TR:040380 Q40380 ARABINGALACTAN-PROTEIN
PRECUSOR. ; mRNA sequence.

ACCESSION A1736967
VERSION A1736967.1 GI:5058435
KEYWORDS EST.

SOURCE soybean.
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 422)

REFERENCE

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Coryell, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)

TITLE

JOURNAL

Public Soybean EST Project

Contact: Shoemaker R/Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

FEATURES

source

Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert Length: 530 Std Error: 0.00
Seq primer: -40RP from gibco
High quality sequence stop: 409.
Location/Qualifiers
1.422
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1013-27"
/clone_lib="Gm-c1013"
/tissue_type="Whole seedlings, 2-3 week old seedlings,
greenhouse grown"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from whole seedlings of 2-3 week old greenhouse grown
plants. The cDNA library was prepared using the Stratagene
pBluescript II XR cDNA library construction kit.
Complementary DNA was synthesized from mRNA using a primer
consisting of a poly (dT) sequence with a XhoI restriction
site. EcoRI adapters were ligated to the blunt-ended cDNA
fragments followed by XhoI digestion. The cDNA fragments
were directionally cloned into the EcoRI-XhoI restriction
site of the pBluescript vector. The ligated cDNA fragments
were transformed into XL10-Gold host cells. This library
was constructed by Dr. Randy Shoemaker and Dr. John
Erpelidg."

BASE COUNT 83 a 105 c 91 g 138 t 5 others
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 422;
Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgagctctgctg 16
|||||

Db 183 gttgctgagctctgctg 198

RESULT 20

LOCUS AM704839 434 bp mRNA EST 21-NOV-2000
DEFINITION SK40B11.Y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1019-5014 5' mRNA sequence.

ACCESSION AM704839
VERSION AM704839.1 GI:7589057
KEYWORDS EST.

SOURCE soybean.
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 434)

REFERENCE

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Coryell, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)

TITLE

JOURNAL

Public Soybean EST Project

Contact: Shoemaker R/Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Insert length: 553 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 413.
 Location/Qualifiers
 1..434
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl019-5014"
 /clone_lib="Gm-cl019"
 /tissue_type="Immature seed coats of greenhouse grown
 plants"
 /lab_host="DH10B (Gibco BRL)"
 /note="Vector: pSPORT1 (Life Technologies); Site_1: Not I;
 Site_2: Sal I; This cDNA library was constructed from mRNA
 isolated from immature seed coats (200-300 mgs) of
 greenhouse grown plants. The library was prepared using
 the Life Technologies pSuperScript cDNA library
 construction kit. Complementary DNA was synthesized from
 mRNA using a poly (dT) sequence with a Not I restriction
 site. Sal I linker adapters were ligated to the
 blunt-ended cDNA fragments followed by Not I digestion.
 The cDNA fragments were directionally cloned into the Not
 I-Sal I restriction site of the pSPORT1 vector. The
 ligated cDNA fragments were transformed into E.coli
 Electromax DH10B host cells (Gibco BRL). This library was
 constructed by Dr. Lila Vockin and Dr. Anu Khanna."

BASE COUNT 57 a 169 c 90 g 118 t
 ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 434;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtgtcgtgacgcctg 16
 ||||||||||||
 Db 398 GTTGGTGGATCTGCTG 413

RESULT 21
 A1293622 436 bp mRNA EST 19-APR-2001
 LOCUS LP06802.5prime LP Drosophila melanogaster larval-early pupal POT2
 DEFINITION Drosophila melanogaster cDNA clone LP06802.5prime, mRNA sequence.
 ACCESSION A1293622
 VERSION A1293622.1 GI:3943029
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 436)
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
 Lewis, S. and Rubin, G.M.
 BDDP/HMT Drosophila EST Project
 Unpublished (2001)
 TITLE JOURNAL
 COMMENT BDDP
 Contact: Stapleton, M.
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, estfruitfly.berkeley.edu
 Place: 68 row: A column: 2
 High quality sequence stop: 390.
 Location/Qualifiers

FEATURES

source 1..436
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="LP06802"
 /clone_lib="LP Drosophila melanogaster larval-early pupal
 POT2"
 /sex="male and female"
 /dev_stage="larvae-pupae"
 /lab_host="DH5-alpha"
 /note="Organ: whole body; Vector: POT2; Site_1: EcoRI;
 Site_2: XhoI; Sized fractionated cDNAs were directly
 ligated into POT2. Plasmid cDNA library."
 BASE COUNT 111 a 122 c 111 g 92 t
 ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 436;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtgtcgtgacgcctg 16
 ||||||||||||
 Db 328 GTTGGTGGATCTGCTG 313

RESULT 22
 BE190258 441 bp mRNA EST 22-JUN-2000
 LOCUS SO09108.y1 Gm-cl035 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-cl035-2848 5', mRNA sequence.
 ACCESSION BE190258
 VERSION BE190258.1 GI:8669151
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae;
 Glycine.
 1 (bases 1 to 441)
 Shoemaker, R., Kelm, P., Vockin, L., Erpelting, J., Corryell, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterson, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 440.
 Location/Qualifiers

FEATURES
 source

1..441
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl035-2848"
 /clone_lib="Gm-cl035"
 /tissue_type="Immature leaves of greenhouse grown plants"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
 cDNA library was constructed from mRNA isolated from
 immature leaves (unfurling trifoliolate) of greenhouse grown

plants that were 2 weeks old. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a NotI restriction site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

BASE COUNT 102 a 63 c 85 g 190 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 441;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gttgctgactgctg 16
|||||
Db 48 GTTGTGATCTGCTG 63

RESULT 23

AL373871

LOCUS

AL373871 470 bp mRNA EST 03-AUG-2000

DEFINITION MEB03D01F1 MEBB Medicago truncatula cDNA clone MEB03D01 T3, mRNA

ACCESSION

AL373871

VERSION

AL373871.1

KEYWORDS

GI:9673623

SOURCE

EST

ORGANISM

barrel medic.

REFERENCE

Medicago truncatula

AUTHORS

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

1 (bases 1 to 470)

Jouneet,E.P., Crespeau,H., van Tuinen,D., Gouzy,J., Jallion,O., Niebel,A., Carreau,Y., Chatagnier,O., Kahn,D., Glaninazzi-Pearson,V. and Gamas,P.

TITLE

Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules

COMMENT

Unpublished (2000)

JOURNAL

Genoscope - Centre National de Sequencage

COMMENT

BP 191 91006 Evry cedex - France

CONTACT

Email: seque@genoscope.cns.fr, Web: www.genoscope.cns.fr

CONTACT

Biologie Molculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email: ht-est@genoscope.cns.fr)

LOCATION

http://sequence.toulouse.inra.fr/Mtruncatula.html).

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

LOCATION

1..470

BASE COUNT

85 a 137 c 68 g 180 t

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 470;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gttgctgactgctg 16
|||||
Db 249 GTTGTGATCTGCTG 264

RESULT 24

AM704996

474 bp mRNA EST 18-APR-2000

LOCUS

SK41C02.Y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1019-5115 5', mRNA sequence.

DEFINITION

AM704996

ACCESSION

AM704996.1

VERSION

GI:7569217

KEYWORDS

EST

SOURCE

soybean.

ORGANISM

Glycine max

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

1 (bases 1 to 474)

Shoemaker,R., Kelm,P., Vodkin,L., Erpelidg,J., Corvett,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kueba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Schaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

TITLE

Public Soybean EST Project

JOURNAL

Unpublished (1999)

COMMENT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

CONTACT

Public Soybean EST Project

site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORI1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lilia Vodkin and Dr. Anu Khanna."

BASE COUNT 66 a 175 c 104 g 129 t

Query Match 64.0%; Score 16; DB 10; Length 474;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgtctgagatctgctg 16
|||||
Db 398 GTTGCTGATCTGCTG 413

RESULT 25
AL379973 481 bp mRNA 03-AUG-2000
LOCUS MTBB48E08F1 MTBB Medicago truncatula cDNA clone MTBB48E08 T3, mRNA
DEFINITION
ACCESSION AL379973
VERSION AL379973.1 GI:9679725
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 481)
Journel,E.P., Crespeau,H., van Tuinen,D., Gouzy,J., Jaillon,O.,
Niebel,A., Carreau,V., Chataigner,O., Kahn,D., Glanmazzi-Pearson
V. and Gamas,P.
Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
nodules
Unpublished (2000)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journel, Laboratoire de
Biologie Moleculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
Mt-est@toulouse.inra.fr Website :
http://sequence.toulouse.inra.fr/Mtruncatula.html).

FEATURES

source
1..481
Location/Qualifiers
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MTBB48E08"
/clone_lib="MTBB"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with
Sinorhizobium meliloti"
/note="Vector: pBluescript PSK; Site_1: EcoRI; Site_2:
XhoI; Plants were grown in an aeroponic chamber on
nitrogen-rich medium for 21 days. Three days before
inoculation with Sinorhizobium meliloti, the medium was
replaced by N-free medium. Root nodules (+ short adjacent
root segments) were harvested 4 days post inoculation.
cDNA was prepared from polyA+ enriched RNA. The cDNA was
directionally ligated into Uni-zap XR vector from
stragene and packaged using GigaPack Gold packaging
extracts. Plasmids containing cDNA inserts were
mass-excised from phage stocks using Exsist helper phage
and propagated in SOLR cells. Clone ordering and
sequencing was performed by the Centre National de

Sequencage (Genoscope, Evry, France)."

BASE COUNT 89 a 152 c 64 g 176 t

Query Match 64.0%; Score 16; DB 10; Length 481;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgtctgagatctgctg 16
|||||
Db 385 GTTGCTGATCTGCTG 400

RESULT 26
BF461669 497 bp mRNA EST 04-DEC-2000
LOCUS BF461669/c
DEFINITION
ACCESSION BF461669
VERSION BF461669.1 GI:11530852
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 497)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL MEDLINE
AUTHORS
TITLE
COMMENT
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
Oligo-dt track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENERICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=No.

FEATURES

source
1..497
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-CG0P-dmu-g-12-0-UI"
/clone_lib="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pTZ19-D-pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_Ret4_S2 library is a subtracted library,
ultimately derived from mouse retina tissue libraries at
various stages of development. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.utoowa.edu.
TAG_SEQ=None found"

BASE COUNT 112 a 168 c 171 g 44 t 2 others

Query Match 64.0%; Score 16; DB 11; Length 497;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgtctgagatctgctg 16

KEYWORDS EST.
SOURCE Lotus japonicus.
ORGANISM Lotus japonicus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
 Lotus.
REFERENCE 1 (bases 1 to 513)
 Colebatch, G., Freund, S., Treweek, B and Udvardi, M.
AUTHORS Lotus japonicus root module ESTs: tools for functional genomics
TITLE Unpublished (2000)
JOURNAL Contact: Udvardi, MK
COMMENT Molecular Plant Nutrition
 Max Planck Institute of Molecular Plant Physiology
 Am Muehlenberg 1, 14476 Golm, Germany
 Fax: 49 331 567 8250
 Email: udvardi@mpimp-golm.mpg.de
 Seq primer: 77
 High quality sequence stop: 513.
FEATURES Location/Qualifiers
 source
 1..513
 /organism="Lotus japonicus"
 /cultivar="Gifu (B-129)"
 /db_xref="taxon:34305"
 /clone_lib="Lotus japonicus node library 5 and 7
 week-old"
 /dev_stage="5 and 7 week-old plants"
 /note="Organ: Node; Vector: pSPORT1; Site_1: Salt;
 Site_2: NoCl; The library was prepared using mRNA
 extracted from nodules of 5 and 7 week-old Lotus plants.
 Nodules were induced by, and contained Mesorhizobium
 strain R7A."
 BASE COUNT 85 a 164 c 99 g 165 t
ORIGIN
 Query Match 64.0%; Score 16; DB 11; Length 513;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 392 GTTGCTGATCTGCTG 407
 QY 1 gttgctgctgctgctg 16
 ||||||||||||||||
 Db 392 GTTGCTGATCTGCTG 407
RESULT 30
LOCUS AL373872 517 bp mRNA EST 03-AUG-2000
DEFINITION MCB03D01R1 MBB Medicago truncatula cDNA clone MBB03D01 T7, mRNA
 sequence.
ACCESSION AL373872
VERSION AL373872.1 GI:9673624
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
REFERENCE 1 (bases 1 to 517)
 Journet, E.P., Crespeau, H., van Tuinen, D., Gouzy, J., Jallion, O.,
 Niebel, A., Carreau, V., Chatagnier, O., Kahn, D., Gianinazzi-Pearson
 V., and Gamas, P.
AUTHORS Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
 nodules
TITLE Unpublished (2000)
JOURNAL Contact: Genoscope
COMMENT Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
 Biologie Moleculaire des Relations Plantes-Microorganismes,
 CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :

MT-est@toulouse.inra.fr Website :
 http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES Location/Qualifiers
 source
 1..517
 /organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"
 /clone_lib="MCB03D01"
 /clone_lib="MTB"
 /tissue_type="symbiotic root nodules"
 /dev_stage="harvested 4 days post inoculation with
 Sinorhizobium meliloti"
 /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
 XhoI; Plants were grown in an aeroponic chamber on
 nitrogen-rich medium for 21 days. Three days before
 inoculation with Sinorhizobium meliloti, the medium was
 replaced by N-free medium. Root nodules (+ short adjacent
 root segments) were harvested 4 days post inoculation.
 cDNA was prepared from polyA+ enriched RNA. The cDNA was
 directionally ligated into Uni-Zap XR vector from
 Stratagene and packaged using Gigapack Gold packaging
 extracts. Plasmids containing cDNA inserts were
 mass-excised from phage stocks using ExAssit helper phage
 and propagated in SOLR cells. Clone ordering and
 sequencing was performed by the Centre National de
 Sequencage (Genoscope, Evry, France)."
BASE COUNT 104 a 147 c 71 g 195 t
ORIGIN
 Query Match 64.0%; Score 16; DB 10; Length 517;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 274 GTTGCTGATCTGCTG 289
 QY 1 gttgctgctgctgctg 16
 ||||||||||||||||
 Db 274 GTTGCTGATCTGCTG 289
RESULT 31
LOCUS A0692054 517 bp DNA GSS 06-JUL-1999
DEFINITION HS_5408_A2_E07_T7A RPCI-11 Human Male BAC library Homo sapiens
 genomic clone Plate-984 Col-14 Row-1, DNA sequence.
ACCESSION A0692054
VERSION A0692054.1 GI:5382302
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 517)
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
 Hood, L.
AUTHORS Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
TITLE Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL Contact: Mahairas GG, Wallace JC, Hood L
COMMENT High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
 or from Research Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 984 row: I column: 14

Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 517.
 Location/Qualifiers
 1. 517

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RPEC-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
 BASE COUNT 154 a 96 c 83 g 178 t 6 others
 ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 517;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 atctgctgttgagc 24
 ||||||||||||
 Db 411 ATCTGCTGTTCAGC 426

RESULT 32
 BFA19551/c 521 bp mRNA EST 28-NOV-2000
 LOCUS UI-R-CA0-Bpd-c-11-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
 DEFINITION UI-R-CA0-Bpd-c-11-0-UI 3', mRNA sequence.
 ACCESSION BFA19551
 VERSION BFA19551.1 GI:11407540
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 521)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL 97044477
 MEDLINE
 COMMENT

CONTACT: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Oligo-dT track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source

Location/Qualifiers
 1. 521
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone_lib="UI-R-CA0-Bpd-c-11-0-UI"
 /clone_lib="UI-R-CA0"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
 library is a subtracted library derived from the following
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons
 , midbrain, cerebral cortex, corpus striatum, testis, and

hippocampus. For a detailed description of the library
 from which this clone was derived, please visit our web
 site at ratest.eng.uiowa.edu. The subtraction has been
 previously described in (Bonaldo, Lennon and Soares,
 Genome Research 6:791-806, 1996)
 TAG-Seq=None found"
 BASE COUNT 113 a 179 c 183 g 46 t
 ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 521;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gtgtgatgctgctg 16
 ||||||||||||
 Db 246 GTTGCTGATCTGCTG 231

RESULT 33
 A1854720/c 522 bp mRNA EST 15-JUL-1999
 LOCUS UI-R-BH0-akc-f-12-0-UI.s1 NIH-BMAP_M.S1 Mus musculus cDNA clone
 DEFINITION UI-R-BH0-akc-f-12-0-UI 3', mRNA sequence.
 ACCESSION A1854720
 VERSION A1854720.1 GI:5498626
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 522)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL 97044477
 MEDLINE
 COMMENT

CONTACT: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: MEST@mail.nih.gov
 Oligo-dT track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
 available by the means that is soon to be determined. When NIH
 determines the means for distribution of the BMAP cDNA clones, this
 record will be updated accordingly when that means is determined.
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source

Location/Qualifiers
 1. 522
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="UI-M-BH0-akc-f-12-0-UI"
 /clone_lib="NIH-BMAP_M.S1"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; The
 NIH-BMAP_M.S1 library is a subtracted library derived from
 a mixture of normalized libraries from ten regions of the
 mouse brain (cerebellum, brain stems, olfactory bulbs,
 hypothalamus, cortex, amygdala, basal ganglia, pineal
 gland, striatum, hippocampus). The driver used for
 subtraction consisted of a pool of 20,000 cDNA clones
 obtained from non-normalized and normalized libraries of
 these ten regions of the mouse brain.
 TAG_LIB=NIH-BMAP_M.S1
 TAG_TISSUE=hypothalamus

BASE COUNT 118 a 174 c 179 g 51 t
ORIGIN TAG-SEQ=CGGTA"

Query Match 64.0%; Score 16; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtctctgagctctg 16
|||||
Db 245 GTTCTGATCTGCTG 230

RESULT 34
AL379974 525 bp mRNA EST 03-AUG-2000
LOCUS MEBB48E08R1 MBB Medicago truncatula cDNA clone MBB48E08 T7, mRNA
DEFINITION sequence.
ACCESSION AL379974
VERSION AL379974.1 GI:9679726
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 525)

REFERENCE
Journet,E.P., Crespeau,H., Van-Tuinen,D., Gouzy,J., Jallou,O.,
Nebel,A., Carreau,V., Chataigner,O., Kahn,D., Gianinazzi-Pearson
, V. and Gamas,P.
Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
nodules
unpublished (2000)

JOURNAL
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, laboratoire de
Biologie Moleculaire des Relations Plantes-Microorganismes,
CNRS-IRRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
Mt-est@toulouse.inra.fr Website :
http://sequence.toulouse.inra.fr/Mttruncatula.html).
Location/Qualifiers

FEATURES
source
1..525
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MEBB48E08"
/clone_1lb="MBB"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with
Sinorhizobium meliloti"
/note="Vector: pBluescript PSK; Site_1: EcoRI; Site_2:
XhoI; Plants were grown in an aerobic chamber on
nitrogen-rich medium for 21 days. Three days before
inoculation with Sinorhizobium meliloti, the medium was
replaced by N-free medium. Root nodules (+ short adjacent
root segments) were harvested 4 days post inoculation.
cDNA was prepared from polyA+ enriched RNA. The cDNA was
directionally ligated into Uni-zap XR vector from
Stratagene and packaged using Gigapack Gold packaging
extracts. Plasmids containing cDNA inserts were
mass-excised from phage stocks using Exsacit helper phage
and propagated in SOLR cells. Clone ordering and
sequencing was performed by the Centre National de
Sequencage (Genoscope, Evry, France)."

BASE COUNT 108 a 148 c 81 g 188 t
ORIGIN

Query Match

64.0%; Score 16; DB 10; Length 525;

Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtctctgagctctg 16
|||||
Db 255 GTTCTGATCTGCTG 270

RESULT 35
AW395303 531 bp mRNA EST 17-JUL-2000
LOCUS sh46d09.y1 Gm-cl017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl017-4746 5' similar to SW:RPB9.DROME P36958 DNA-DIRECTED RNA
POLYMERASE II 15.1 KD POLYPEPTIDE ;, mRNA sequence.
ACCESSION AW395303
VERSION AW395303.1 GI:6913773
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 531)

REFERENCE
AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Corryell,V., Khanna
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Willson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert length: 743 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 342.
Location/Qualifiers

FEATURES
source
1..531
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl017-4746"
/clone_1lb="Gm-cl017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="XJ10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from vegetable buds of field grown plants. The cDNA
library was prepared using the Stratagene pBluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adaptors
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into XJ10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelting."

BASE COUNT 159 a 96 c 107 g 169 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 531;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtctgcatctctcgc 16
 |||
 Db 60 GTTCTGATCTGCTG 75

RESULT 36
 A0889100/c 555 bp DNA GSS 10-NOV-1999
 DEFINITION HS_2204_AL_C07_MR_C1T Approved Human Genomic Sperm Library D Homo
 A0889100 sapiens genomic clone Plate=2204 Col=13 Row=E, DNA sequence.
 ACCESSION A0889100.1 GI:6345290
 VERSION GSS.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 555)
 AUTHORS Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahaitas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2204 row: E column: 13
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 555.

FEATURES
 source Location/Qualifiers
 1..555
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2204 Col=13 Row=E"
 /clone_lib="C1T Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
 E-Coli DH10B"

BASE COUNT 131 a 135 c 147 g 124 t 18 others

Query Match 64.0%; Score 16; DB 13; Length 555;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctgctgtttgaacg 25
 |||
 Db 497 TCTGCTGTTGAGCG 482

RESULT 37
 B1229421/c 570 bp mRNA EST 11-JUL-2001
 LOCUS B1229421
 DEFINITION RE27194.5 prime RE Drosophila melanogaster normalized Embryo pF1c-1
 Drosophila melanogaster cDNA clone RE27194.5 similar to CG1148;
 FBan0001148 located on: 3R 83E1-83E1.1; 04/12/2001, mRNA sequence.
 ACCESSION B1229421
 VERSION B1229421.1 GI:14696685
 KEYWORDS EST.

SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 570)
 AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
 J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
 Mungall,C.J., Nuno,J., Pacled,J., Paragas,V., Park,S.,
 Phonsavanong,S., Wan,K., Yu,C., Lewis,S.E., Celinker,S. and Rubin
 G.M.
 BDGP/HMI RE Drosophila EST Project
 UNPUBLISHED (2001)
 CONTACT: Stapleton, M.
 BDGP

TITLE Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
 hit genomic AB003600; arm:3R [1868408,2159552]
 estimated-cyto:83D2-83E4; 04/12/2001
 plate: RE:271 row: H column: 10
 High quality sequence stop: 545.

FEATURES
 source Location/Qualifiers
 1..570
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="RE27194"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pF1c-1"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DH5-alpha TonaA"
 /note="Organ: embryo; Vector: pF1c1; Site_1: XhoI; Site_2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

BASE COUNT 139 a 179 c 153 g 99 t

Query Match 64.0%; Score 16; DB 11; Length 570;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtctgcatctctcgc 16
 |||
 Db 63 GTTCTGATCTGCTG 48

RESULT 38
 AW573807 574 bp mRNA EST 07-SEP-2000
 LOCUS AW573807
 DEFINITION EST316398 GVN Medicago truncatula cDNA clone pgvW-48022, mRNA
 sequence.
 ACCESSION AW573807
 VERSION AW573807.1 GI:7238540
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 REFERENCE 1 (bases 1 to 574)
 AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gant,J.S., Peng
 H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
 Holt,I.E. and Fraser,C.M.
 ESTs from one month old nitrogen-fixing root nodules of Medicago
 truncatula
 UNPUBLISHED (2000)
 CONTACT: Carroll P. Vance

Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name: M252705e
TIGR sequence name: M2CAF957K
More information is available at:
'http://chryslie.tamu.edu/medicago'
Seq primer: SKnod (CTA gAA CTA gTg gAT CC).

FEATURES

source

1. 574
Location/Qualifiers
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pgvsn-48022"
/clone_lib="GSN"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the Uni-ZAP XR vector from StrataGene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells."

BASE COUNT

110 a 165 c 80 g 218 t

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 574;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtgtctgacatctg 16
|||||
Db 370 GTTGCTGATCTGCTG 385

RESULT 39

BE999070

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

1 (bases 1 to 575)

Fedorova, M., Pierson, B. L., Samac, D. A., Gantt, J. S., Vance, C. P., Town, C. D., Bowman, C. L., Craven, M. B., Cho, J., and Fraser, C. M.

ESTs from senescent nodules of Medicago truncatula

Unpublished (2000)

Contact: Carroll P. Vance

Department of Agronomy and Plant Genetics

University of Minnesota

411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA

Tel: 612 625 5715

Fax: 651-649-5058

Email: vance004@maroon.tc.umn.edu

University of Minnesota name: M273048e TIGR sequence name:

M273048e TIGR sequence name:

http://chryslie.tamu.edu/medicago

Seq primer: SKnod (CTA gAA CTA gTg gAT CC).
Location/Qualifiers
1. 575
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pgvsn-15C19"
/clone_lib="GSN"
/tissue_type="senescent root nodules"
/dev_stage="mixture of effective nodules from 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage"
/lab_host="E. coli strain SOLR"
/note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of effective nodules of 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage. The cDNA was directionally ligated into the Uni-ZAP XR vector from StrataGene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

FEATURES

source

BASE COUNT 103 a 164 c 82 g 226 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 575;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtgtctgacatctg 16
|||||
Db 347 GTTGCTGATCTGCTG 362

RESULT 40

BF525150

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

REFERENCE

1 (bases 1 to 580)

Bonaldi, M. F., Lennon, G., and Soares, M. B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

cDNA library Preparation: M. B. Soares lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

This clone is also available through the I.M.A.G.E. Consortium at

LNL (info@imge.llnl.gov). IMAGE ID= 1795307

Seq primer: M13 Forward.

Location/Qualifiers

1. 580

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-AF0-yd-e-05-0-UI"

/clone_lib="UI-R-AF0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-AF0
library is a non-normalized library constructed from 15
dpc rat atriocentricular (AV) canal. The tag is a string
of 5 nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
1996. Tissue provided by Jim Lin, Department of Biology,
University of Iowa."

BASE COUNT 155 a 146 c 147 g 132 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 580;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcctgacatcgtctgt 18
|||||
Db 321 TCGTCGATCTGCTGTT 336

RESULT 41

AM734894

LOCUS 591 bp mRNA EST 24-APR-2000
DEFINITION sk76d03.y1 Gm-cl016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl016-10014 5', mRNA sequence.

ACCESSION

AM734894

AM734894.1 GI:7640531

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE

AUTHORS

1 (bases 1 to 591)

Shoemaker, R., Kelm, P., Vodka, L., Erpelting, J., Coryell, V., Khanna,
A., Bolla, B., Marra, M., Hüller, L., Kucaba, T., Martin, J., Beck, C.,
Wyllie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers,
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann,
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 380.

TITLE

JOURNAL

COMMENT

Location/Qualifiers

1. 591

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl016-10014"

/clone_lib="Gm-cl016"

/tissue_type="immature flowers of field grown plants"

/lab_host="XL10-Gold"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-AF0
library is a non-normalized library constructed from 15
dpc rat atriocentricular (AV) canal. The tag is a string
of 5 nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
1996. Tissue provided by Jim Lin, Department of Biology,
University of Iowa."XhoI; This cDNA library was constructed from mRNA isolated
from immature flowers of field grown plants. The cDNA

FEATURES

source

Location/Qualifiers

1. 591

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl016-10014"

/clone_lib="Gm-cl016"

/tissue_type="immature flowers of field grown plants"

/lab_host="XL10-Gold"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-AF0
library is a non-normalized library constructed from 15
dpc rat atriocentricular (AV) canal. The tag is a string
of 5 nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
1996. Tissue provided by Jim Lin, Department of Biology,
University of Iowa."XhoI; This cDNA library was constructed from mRNA isolated
from immature flowers of field grown plants. The cDNA

library was prepared using the Strategene pluscript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pluscript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelting."

BASE COUNT 121 a 180 c 103 g 185 t 2 others
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 591;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttcctggatcgtctgt 16
|||||
Db 330 GTTCTGATCTGCTGCTG 345

RESULT 42

AZ267971

LOCUS 601 bp DNA GSS 26-JUL-2000
DEFINITION RPCI-23-103L11.TJ RPCI-23 Mus musculus genomic clone RPCI-23-103L11
DNA sequence.

ACCESSION

AZ267971

AZ267971.1 GI:9481584

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 601)

REFERENCE

AUTHORS

Zhao, S., Nieman, W., Feidby, T., Malek, J., Shatman, S., Ahnert,
B., Levins, M., Moggan, S., Tsegaye, G., Geer, K., Kroll, M., de Jong, P.
and Fraser, C. M.
Mouse BAC end sequences from library RPCI-23
Unpublished (1999)
Other GSS: RPCI-23-103L11.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buhalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buhalo.edu/orderingframe.htm>)
or from Resea ch Genetics (<http://resgen.com>). BAC end page:
http://www.tigr.org/tbcbac/ends/mouse/bac_end_intro.html
Plate: 103 row: L column: 11
Seq primer: Sp6
Class: BAC ends.

TITLE

JOURNAL

COMMENT

Location/Qualifiers

1. 601

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-103L11"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the

FEATURES

source

Location/Qualifiers

1. 601

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-103L11"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the

FEATURES

source

Location/Qualifiers

1. 601

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-103L11"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the

ECORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).
 BASE COUNT 189 a 92 c 111 g 209 t
 ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 601;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtcgtgactcgtcgt 17
 |||
 Db 354 ttgtcgtgactcgtcgt 369

RESULT 43
 BI229406/c 602 bp mRNA EST 11-JUN-2001
 LOCUS RE27178.5prline RE Drosophila melanogaster normalized Embryo pfic-1
 DEFINITION Drosophila melanogaster cDNA clone RE27178.5 similar to CG1148:
 FBan0001148 located on: 3R 83E1-83E1.; 04/12/2001, mRNA sequence.

ACCESSION BI229406
 VERSION BI229406.1 GI:14696670
 KEYWORDS EST
 SOURCE fruit fly

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 602)
 Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nuno, J., Pacleb, J., Paragov, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Robin, G.M.

AUTHORS

TITLE BDCP/HMMI RE Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BDCP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est/fruitfly_berkeley.edu
 hit genomic AE003600: arm:3R [1868408..2159952]

Estimated cyto:83D2-83E4: 04/12/2001

Plate: RE.271 row: G column: 6

High quality sequence stop: 476.

Location/Qualifiers

FEATURES

SOURCE

1..602

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="RE27178"

/clone_lib="RE Drosophila melanogaster normalized Embryo pfic-1"

/sex="male and female"

/dev_stage="0-24 hours mixed stage embryonic"

/lab_host="DH5-alpha Tona"

/note="Organ: embryo; Vector: pfic1; Site: 1; XhoI: Site: 2; BamHI: Library was kindly generated by Piero Carlucci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

BASE COUNT 148 a 185 c 161 g 108 t
 ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 602;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgactcgtcgt 16
 |||
 Db 63 gtgtcgtgactcgtcgt 48

RESULT 44
 AM267738
 LOCUS

DEFINITION AM267738 609 bp mRNA EST 07-SEP-2000
 EST305866 DSTR Medicago truncatula cDNA clone pDSIR-717, mRNA
 sequence.

ACCESSION AM267738
 VERSION AM267738.1 GI:6654694
 KEYWORDS EST

SOURCE barrel medic.
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

REFERENCE 1 (bases 1 to 609)
 Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, G.S., Peng, H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E., and Fraser, C.M.

ESTs from roots of Medicago truncatula after inoculation with Phytophthora medicaginis
 Unpublished (1999)
 CONTACT: Carroll P. Vance

Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 191 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058

Email: vance004@maroon.tc.umn.edu
 Minnesota EST name: M250222e
 TRIR sequence name: M250222e

More information, including clone ordering, is available at.

<http://chrystle.tamu.edu/medicago>

Seq primer: SKmod (CTA gaa gta gta gta CC).
 Location/Qualifiers

FEATURES

SOURCE

1..609

/organism="Medicago truncatula"

/cultivar="genotype A17"

/db_xref="taxon:3880"

/clone="pDSIR-717"

/clone_lib="DSIR"

/tissue_type="roots infected with Phytophthora medicaginis"

/dev_stage="roots harvested at 10 days post inoculation with Phytophthora medicaginis"

/lab_host="E. coli strain XLOLR"

/note="Vector: pBluescript SK-; Site: 1: EcorI; Site: 2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10 days post inoculation with Phytophthora medicaginis. The cDNA was directionally ligated into the uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells. Note: EST may be of fungal origin."

BASE COUNT 117 a 172 c 82 g 238 t

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 609;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgactcgtcgt 16
 |||
 Db 381 gtgtcgtgactcgtcgt 396

RESULT 45
 AM980891
 LOCUS AM980891 616 bp mRNA EST 07-SEP-2000

DEFINITION EST392044 GVN Medicago truncatula cDNA clone pGVN-60C11, mRNA
sequence.

ACCESSION AM980891

VERSION AM980891.1 GI:8172436

KEYWORDS EST.

SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales; Medicago.

REFERENCE 1 (bases 1 to 616)
Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gault,J.S., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula

JOURNAL Unpublished (2000)

COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@amaroon.tc.umn.edu
University of Minnesota name:M256725e
TIGR sequence name:MTCBW187K
More information is available at: <http://chryslie.tamu.edu/medicago>
Seq primer: SKmod (CTA GAA CTA gta gta CC).
Location/Qualifiers

FEATURES
source
1..616
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVN-60C11"
/clone_lib="GVN"
/issue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XL0LR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XL0LR cells."

BASE COUNT 118 a 173 c 82 g 243 t

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 616;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgattctgctg 16
|||||

Db 388 gttgctgattctgctg 403

Search completed: December 26, 2001, 14:10:04
Job time: 7051 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:50:57 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196F-4

Perfect score: 25

Sequence: 1 ggaatcgtctgttgaagcgacgacg 25

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 252155

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :
1: GenEmbl:*
2: gb_ba:*
3: gb_hlg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_pat:*
8: gb_ph:*
9: gb_pl:*
10: gb_pr:*
11: gb_ro:*
12: gb_sts:*
13: gb_sy:*
14: gb_un:*
15: gb_vi:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_sy:*
29: em_un:*
30: em_vi:*
31: em_hlgo_hum:*
32: em_hlgo_inv:*
33: em_hlgo_rnd:*
34: em_hlgo_hum:*
35: em_hlgo_inv:*
36: em_hlgo_rnd:*
37: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	6 AR029499	AR029499 Sequence
2	25	100.0	1041	6 AR034916	AR034916 Sequence
3	25	100.0	1084	6 A11530	A11530 B10b gene o
4	25	100.0	5526	1 AF250776	AF250776 Unculture
5	25	100.0	5793	1 ECOBIO	J04423 E.coli 7/8-
6	25	100.0	5872	6 A38246	A38246 Sequence 1
7	25	100.0	5872	6 A38251	A38251 Sequence 1
8	25	100.0	5872	6 A33674	A33674 Sequence 1
9	25	100.0	5872	6 A33679	A33679 Sequence 6
10	25	100.0	5872	6 AR101809	AR101809 Sequence
11	25	100.0	5872	6 AR101810	AR101810 Sequence
12	25	100.0	11022	1 AE000180	AE000180 Escherich
13	25	100.0	11022	1 AE004192	AE004192 Vibrio ch
14	25	100.0	13891	1 AE005258	AE005258 Escherich
15	25	100.0	13501	1 AP002553	AP002553 Escherich
16	21	84.0	128	1 ECOBIOB	M2731 Escherichia
17	21	84.0	8227	1 AF248314	AF248314 Unculture
18	18	72.0	1041	1 EMBIOTOP2	U38648 Erwinia her
19	18	72.0	1121	6 E00893	E00893 Genomic DNA
20	17	68.0	1044	5 AF250347	AF250347 Xenopus 1
21	17	68.0	11614	5 AF281080	AF281080 Xenopus 1
22	17	68.0	14268	1 AE001220	AE001220 Treponema
23	16	64.0	801	6 AX120335	AX120335 Sequence
24	16	64.0	924	6 AX064067	AX064067 Sequence
25	16	64.0	6887	8 YL1012084	YL1012084 Yarrowia
26	16	64.0	11823	1 AE002054	AE002054 Deinococc
27	16	64.0	13856	1 AE004026	AE004026 Xylella f
28	16	64.0	40031	3 CBRG0109	U56248 Caenorhabdi
29	16	64.0	156425	8 AP002897	AP002897 Oryza sat
30	16	64.0	163489	7 AC016697	AC016697 Homo sapi
31	16	64.0	168177	2 AC023914	AC023914 Homo sapi
32	16	64.0	171069	2 AP003845	AP003845 Oryza sat
33	16	64.0	191677	2 AC090534	AC090534 Mus muscu
34	16	64.0	200000	2 AC007104	AC007104 Homo sapi
35	16	64.0	349980	6 AX120085	AX120085 Sequence
36	15	60.0	306	14 PPHE7D	D50548 Human papil
37	15	60.0	625	8 PEU244276	AJ244276 Pringshei
38	15	60.0	810	6 AX110623	AX110623 Sequence
39	15	60.0	1551	1 AF128956	AF128956 Escherich
40	15	60.0	1572	1 SMAAC3VB	M97172 Serratia ma
41	15	60.0	1616	10 AF239157	AF239157 Rattus no
42	15	60.0	1623	10 AF009246	AF009246 Mus muscu
43	15	60.0	2795	1 AF039572	AF039572 Serratia
44	15	60.0	6043	15 PMO238607	AJ238607 Potato mo
45	15	60.0	7779	14 HPU31779	AJ31779 Human papil
46	15	60.0	8219	7 AF066865	AF066865 Bacteriop
47	15	60.0	10881	1 AE004129	AE004129 Vibrio ch
48	15	60.0	11062	1 AE006532	AE006532 Streptoco
49	15	60.0	11180	1 AE007575	AE007575 Clostridi
50	15	60.0	11328	1 AE004322	AE004322 Vibrio ch
51	15	60.0	11467	1 AE006281	AE006281 Lactococc
52	15	60.0	11674	1 AE005391	AE005391 Escherich
53	15	60.0	11684	1 AE000263	AE000263 Escherich
54	15	60.0	13823	1 D90812	D90812 E.coli geno
55	15	60.0	19521	1 D90811	D90811 E.coli geno
56	15	60.0	36949	7 AF323670	AF323670 Bacteriop
57	15	60.0	37555	15 U42832	U42832 Caenorhabdi
58	15	60.0	37991	9 AC005101	AC005101 Homo sapi
59	15	60.0	38245	1 AF105019	AF105019 Salmonell
60	15	60.0	40629	2 AC006105	AC006105 Homo sapi
61	15	60.0	42563	8 AB018110	AB018110 Arabidops
62	15	60.0	43439	2 U82208	U82208 Homo sapien
63	15	60.0	43760	2 AC087367	AC087367 Homo sapi
64	15	60.0	43955	2 AC015342	AC015342 Drosophi
65	15	60.0	46387	2 U82212	U82212 Homo sapien
66	15	60.0	60248	9 AL137181	AL137181 Human DNA
67	15	60.0	60248	9 AL1391647	AL1391647 Human DNA
68	15	60.0	62544	8 NC65A3	AL513411 Neurospor
69	15	60.0	63633	9 AL139409	AL139409 Human DNA
70	15	60.0	69918	9 AL441988	AL441988 Human DNA

```

c 71 15 60.0 84107 9 HS248E1 AL023578 Human DNA
c 72 15 60.0 85084 9 AB042297 Homo sapi
c 73 15 60.0 85516 9 AC009493 Homo sapi
c 74 15 60.0 90720 2 AC023590 Homo sapi
c 75 15 60.0 91201 9 HSDA86H19 AL049562 Human DNA
c 76 15 60.0 106408 9 AL512303 Human DNA
c 77 15 60.0 110000 2 LMFICHR3-21 Continuation (22 o
c 78 15 60.0 110000 2 LMFICHR36-30 Continuation (31 o
c 79 15 60.0 110000 2 AC020850-1 Continuation (2 of
c 80 15 60.0 110716 9 AL135918 Human DNA
c 81 15 60.0 125287 9 HSI359A19 AL031729 Human DNA
c 82 15 60.0 128589 2 AC087122 Mus muscu
c 83 15 60.0 129048 2 AL592286 AL592286 Homo sapi
c 84 15 60.0 132592 2 HS3E5 AL022239 Homo sapi
c 85 15 60.0 134137 9 HS3E5 AL590151 Homo sapi
c 86 15 60.0 136930 2 AC026292 Homo sapi
c 87 15 60.0 137733 2 AC027787 Homo sapi
c 88 15 60.0 138097 2 AC027787 Homo sapi
c 89 15 60.0 138097 2 AC027787 Homo sapi
c 90 15 60.0 146349 2 AC007863 AC007863 Trypanoso
c 91 15 60.0 148891 9 AL162274 AL162274 Homo sapi
c 92 15 60.0 148927 2 AC087127 Mus muscu
c 93 15 60.0 150237 2 AL357335 Homo sapi
c 94 15 60.0 151494 2 AL138931 Homo sapi
c 95 15 60.0 152755 2 AC006358 Homo sapi
c 96 15 60.0 154005 2 AC073208 Homo sapi
c 97 15 60.0 157442 2 AC079929 Homo sapi
c 98 15 60.0 161635 2 AC084824 AC084824 Homo sapi
c 99 15 60.0 161635 2 AC083873 AC083873 Homo sapi
c 100 15 60.0 161651 9 HS233K16 AL035406 Human DNA

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JUN-1999.
FEATURES
    source
        1..1041
        /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggaatcgtctgttgaagcgacag 25
    |||
Db 63 GGATCTGCTGTTGAAGCGACGACG 87

```

```

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

```

```

ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1041)
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999.
FEATURES
    source
        1..1041
        /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggaatcgtctgttgaagcgacag 25
    |||
Db 63 GGATCTGCTGTTGAAGCGACGACG 87

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10b gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
JOURNAL Location/Qualifiers
FEATURES
    source
        1..1084
        /organism="Escherichia coli"
        /db_xref="taxon:562"
        /gene="B10b"
        /gene="B10B"
        /codon_start=1
        /transl_table=1
        /protein_id="CAA00965.1"
        /db_xref="GI:490219"
        /db_xref="SWISS-PROT:P12996"
        /translation="MAHRPRWTLSQVTELFERPLDLLLEPAQOVHRQHPDPROVOYST
        LLSIKTACPEDECKYCPQSSRYRTGLAEKLMVEEVLDESARRAKAGSTRGCMGAW
        KNEHEKMPYLEOVGOKAMGLEACMTLGTLSSEQARIANAGLDYVHNHNDTSPF
        VPTNMLTKVKGTPLEADNDVDADFDRTITAVARIMMPTSYVRLSAGREQMNOTOAMC
        FMAGANSIFYGCKLLTTPNEDEDKDQLFRKIGLNQQTAVAVLAGNEQQQRLEQALMT
        PTDDEVYNAAL"
        /protein_id="P12996"
        /db_xref="taxon:562"
        /gene="B10b"
        /gene="B10B"
        /codon_start=1
        /transl_table=1
        /protein_id="CAA00965.1"
        /db_xref="GI:490219"
        /db_xref="SWISS-PROT:P12996"
        /translation="MAHRPRWTLSQVTELFERPLDLLLEPAQOVHRQHPDPROVOYST
        LLSIKTACPEDECKYCPQSSRYRTGLAEKLMVEEVLDESARRAKAGSTRGCMGAW
        KNEHEKMPYLEOVGOKAMGLEACMTLGTLSSEQARIANAGLDYVHNHNDTSPF
        VPTNMLTKVKGTPLEADNDVDADFDRTITAVARIMMPTSYVRLSAGREQMNOTOAMC
        FMAGANSIFYGCKLLTTPNEDEDKDQLFRKIGLNQQTAVAVLAGNEQQQRLEQALMT
        PTDDEVYNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggaatcgtctgttgaagcgacag 25
    |||
Db 86 GGATCTGCTGTTGAAGCGACGACG 110

```

```

RESULT 4
LOCUS AF250776 5526 bp DNA
DEFINITION Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in

```


SOURCE Escherichia coli (strain K-12) DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 5793)
 AUTHORS Otsuka, A.J., Buoncristiani, M.R., Howard, P.K., Flamm, J. and Johnson, O.
 TITLE The Escherichia coli biotin biosynthetic enzyme sequences
 JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
 MEDLINE 89066784
 COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-Nov-1988

FEATURES
 source
 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 complement(98..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=1
 /protein_id="AA023513.1"
 /db_xref="GI:455168"
 /translation="MKLISNDLRDQKLPRIHVGNGYXGDNISPHILAMDDVPACTK
 SFVYCYDPAIPAGSGMWVYVNLPAIDRPIVPOGSGSLVAMPDGLDTRDPCGTG
 YDGAAPKKEITHYIFTYHLDIERIDYDEGASAAVGNFVSHSLASSTVMS"
 complement(633..1925)
 /gene="bioA"
 complement(633..1925)
 /gene="bioA"
 /codon_start=1
 /transl_table=1
 /product="7,8-diamino-pelargonic acid aminotransferase"
 /protein_id="AA023514.1"
 /db_xref="GI:457106"
 /translation="MTTDLAPDQRIHHPYTSMSPLFVYVNSAEGCELLSDGR
 LVDMSSWMAIHGVNHPOLNMAKSOIDAMSHVMGGITTHAPATIELCKILVMSGRN
 ALCEVLADSGVAIVEAMKMLQYQAKGAEKQRIIFRNCHGDTFGAMVCVCPDN
 SMHSLMKGYLPENLFAPAPQSRMDGDERDVGAFRLAAARHETAAVITPEYOGA
 GGRMYHPEMLKIRKICDREGILLIADILATGFGTGTGKLPACHEAELAPDLCTCA
 LTGGMTLSATLTTRVARTISNGBAGCCHMGPTTNGNLAACAASALATIESDQO
 OOVADIEVQRLQIAPARAENKADVIGATIGVETTHPVMAALQKFFVEGWI
 PRKLIYEMPTIILPOLQRLTAANRAVODETECCQ"
 2012..3052
 /gene="bioB"
 2012..3052
 /gene="bioB"
 /codon_start=1
 /transl_table=1
 /product="biotin synthetase"
 /protein_id="AA023515.1"
 /db_xref="GI:145425"
 /translation="MAHPRRTLSGVTELEPEKLLDILFEAOVYRHOHPDDROVST
 LSTIKGAPEDCKYCPQTSRYKGTGLAEIRMEYVLESARAKAAASTRFGMAAM
 KNPHERDMYLEOMGVKAMLENCMTLGTISEQOQLAAGLDVYNHNLSPER
 VNNITTRTYQERLDTLEKVDAGIKVSGGIVGGETVKRAGLLLOLANIPPEPS
 VTINMLVKYKGPLADNDVDVDFRTITAVARIMPTSYVALSGRQMDQOAMC
 PMAGANSIFYGCKLLTPNPEBDKDLQIFRKILGNPQOTAVIACDNEQOQRLQALMT
 PTDIEYNAAL"
 3049..4203
 /gene="bioF"
 3049..4203
 /gene="bioF"
 /codon_start=1
 /transl_table=1
 /product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA023516.1"
 /db_xref="GI:145426"
 /translation="MSWQETNAALDARADALRRRYPVAGAGRWLVDDROYLNF
 LSNDYLGSHHPQIIIRAWOGAEOFGISGGSVSVVICALLEELKEMLGYSVA
 LLFTSGAANAVIAAMAKEDRIADRLSHASLEAASISPSQLRRFRAHNDYTHLAR
 LLASPCQOMVTEGVSMDGSAPLAIEIGVYQOHNGLMDANDAGTGTGIGQSGC

SCMLQKVPKELLVTEGKGVSGAVALCSSVYADYLLQPARHLITSTSMPPAQAOL
 PASLAYIRSDSDAREKALAILTFRAGVODLPETLADSCSALQIPLVNGNSRALOL
 AKLRQOGWVAIRPVPYAGTARLRLTLAAHMODIDRLLEVLHNG"
 4190..4945
 /gene="bioC"
 4190..4945
 /gene="bioC"
 /codon_start=1
 /transl_table=1
 /protein_id="AA023517.1"
 /db_xref="GI:145427"
 /translation="MATVVKQIAAFGRAAAHVEOHADLQROSDALLAMLPORRYT
 HVLADGCGGMSRMRERHAGVATLDSPPMLVQAROKDAADHYLAGDIESIPLARA
 FPDLAGSNIAVOMCGNLSALRELYRVYRPGVVAETTLVCGSIPERQAMQVDERP
 HANRFLPDEITQSLNGVYOHHTOPITLMDADALSARSLKIGATHLLEGROPRL
 TRSOLQRLQALAPQOQGRPLYHLEGLVARE"
 4938..5597
 /gene="bioD"
 4938..5597
 /gene="bioD"
 /codon_start=1
 /transl_table=1
 /product="methionine synthetase"
 /protein_id="AA023518.1"
 /db_xref="GI:145428"
 /translation="MSKRYEVTGTDTEVKTAVSCALLQAKRAGRYTACYKRPVASS
 EKTEPGLNSDALQORNSQLQDYATVPTFAEPSPHISAGQRPTEISLVMSG
 LRALEQADWVLEVGAGGMPPLSDTFEPDMVQBOPLVITYGVYGLGTINAMLA
 OVIOAGLITLGVAVANDVTPRGKHAEMTTLTMTIRRCHEKSPGLQIKQMRQPS
 7"

BASE COUNT 1363 a 1554 c 1631 g 1245 t
 ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 1; Length 5793;
 Best Local Similarity 100.0%; Pred. No. 0 00013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgcgttgaagcgaacg 25
 |||||
 Db 2074 GGATCTGCTGTTGAAGCGAGCAG 2098

RESULT 6
 LOCUS A38246 5872 bp DNA PAT 05-MAR-1997
 DEFINITION Sequence 1 from Patent W09408023.
 ACCESSION A38246
 VERSION A38246.1 GI:2294844
 KEYWORDS
 SOURCE
 ORGANISM Escherichia coli.
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 JOURNAL Patent: WO 9408023-A 1 14-APR-1994;
 LONZA AG (CH)
 COMMENT
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331
 Other publication JP 8501694T 960227.
 location/Qualifiers
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A-15/9"

```

CDS
    /gene="ORF1"
    /5098..5574
    /gene="ORF1"
    /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
    /codon_start=1
    /transl_table=11
    /number=6
    /evidence=experimental
    /product="PROTEIN"
    /protein_id="CA02328.1"
    /db_xref="GI:2294848"
    /translation="MKLISNDLNDGDKLPRHVFGMGCTGDGNISPLAMDVPAGT
    YFVWCYCPDAPATGSGMMHWVYVNLPAIDFVLPQGGSGIYAMPDSVLTTRDFEKT
    KDGAAPRGETHRYLFTVHALDIERIDVDEGASGAMVGFNVHFHSLASITPMS"
    5583..5605
    5583..5644
    /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
    TERMINATOR"
    /number=2
    /evidence=experimental

BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY
1 ggcacgcacgttgaagcgcacgac 25
|||||
Db 179 GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT 7
A38251 5872 bp DNA PAT 05-MAR-1997
LOCUS
DEFINITION
Sequence 6 from Patent WO9408023.
ACCESSION
A38251
VERSION
A38251.1 GI:2294849
KEYWORDS
.
SOURCE
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
1 (bases 1 to 5872)
Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
COMMENT
Other publication PL 3069301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694P 960227.
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PRO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/number="BIOF"
1154..2308
/number="BIOF"
/number="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental

```

```

/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWEEKINAAADARRADALRRRYPAQAGRWLVADDDROYLN-
SSNYGLGISHHPOIIRAWOQGAEOFGIGSGSGSVGVHQALEBELAEMIGSYSA
LLPTSGFANQAVIAAMAKEDRIADRLSHSLLEAASISPSQIRREPAINDTHLAR
LASPCQOQWVTEGVSGMDGSAPIAEIOQYQNGHLMNDAGTGTVEQCHG
SCMLQYKPELLVYTFGRKGVSSAAVLCSSTVADYLLQFARHLISTSMPPAOQAL
RASLAVIRSDGDKARRKLAALITFRAGVODDPTLADSCSAIOLPLVGDNSRALOI
AEKLQOQCVATIRPPIVPAIGARLRLLITAAHEMODIDRLLEVLHGNG"
RBS
3030..3045
/standard_name="B10D RBS"
3043..3753
/gene="B10D"
3043..3753
/gene="B10D"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFVTGDTTEVGKTVASCALLQAKAAGYRTAGYKPVASG
EKTPGGLRNSDALALQNSLSLODAVAVNYPFAPPTSPHITSAGGRPELISVMSAG
LRALQOQDWLVVAGSGWMTPLSDPTTFADWTQQLPVLYVGVKQGINAMILTA
OVIOHAGITLAVGAVANDYTPPKRHAFTMTLTRMIPAPLLGELPWLAEENPENAATGK
YINAEVDASTLCTFTRSL"
BASE COUNT      1318 a      1352 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gqatcgtcgttggaagcgacag 25
|||||
Db 179 GGATCTGCTGTTGAAGCGACGAG 203

RESULT 8
LOCUS      A93674      5872 bp      DNA
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION  A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch O, and Brass J.
TITLE       Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-Oct-1997;
            LONZA AG (CH)
FEATURES
  source
    1..5872
    /organism="Escherichia coli"
    /strain="DSM498"
    /db_xref="taxon:562"
    /clone="PBO30A-15/9"
    1..96
    /function="PROMOTER PTAC"
    /evidence=experimental
    23..28
    /evidence=experimental
    45..49
    /standard_name="PROMOTER PTAC"
    /evidence=experimental
    105..119
    RBS

```

```

gene
117..1157
/gene="B10B"
117..1157
/gene="B10B"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRRRWTLQVTEIPEKPLIDLFEAQVHROHFDPROVST
LISIKTACPEDECKYCPQSSRYKTGLEAEIRLMEVEVLESARAKAAGSTFEFGAM
KNHREDMAYLEOMYGVKAMGLEACMTIGTLESQAOITLHAGIDYNNHLDPSPE
YGNITTRTYOERLDELKVRDAGIKVSCGCIIGETVYKRACILQIALPTPES
VPIMLVYKGTGPIADNDVDYDFDRIITAVARIMPTSVYRLSAGREOMNEQOAMC
PMAGANSIFYGCKLITTPNPERDKDLFRKLGLNPQOTAVLAGDNEQOORLEQALMT
EDYDYNAAL"
RBS
2284..2297
/standard_name="B10C RBS"
2295..3050
/gene="B10C"
2295..3050
/gene="B10C"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MAIVNKOAIAAFGRAAHYEOHADLQROSADALLMLPQRYT
HVLGCGCPGMSHWRERAAQVTAALDSPMLVOAQKDAADHYLAGDIESPLATA
TFEDLAMSNI.AVQWGNLSTALRELYRVVRPKGVVAFYTLVQSLPELQAMQVDEBP
HANRFLPDEIEOSLNGVYOHHTOPIITLWFDALSMRSLKIGATHLHEGRDRTL
TRSOIORLOLAPOQGGRYPLTYHLFLGVITARE"
3742..3752
/standard_name="B10A RBS"
3750..5039
/gene="B10A"
3750..5039
/gene="B10A"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAPDQRIHWPYTSMTSPLPYVYVSAEGGELLISDGR
LVQGMSSWMAAIGHYVNPOLNAAKMSQIDMSHVMEGJITHAPILCKRLVAMPPO
LECVFLADSGVAEVAMKMALOYOAKGASRORFLTPRNGYGDGFGAMSVCDPNS
MHSIMKGYLPENLEFAPAPQSRMDGEMDERMVGFAHMAHREHETAAVIIEPIYGAG
GMRUYHPEWMLKRIKIDOREGILLADEIATGCRGKXIFACEHAIADIDLCGLAL
TGGMILSATLITREVAETISNCEACGFMHGTPTPGNPLACAAANSLATLSSGWOQ
QVADIEYQLEQALAPADAEVADVLAIGVETTHVYNAALOKPFVEQGVWIRP
FGKLIYMPPIYILPQOLRUTAAVNRVADDEFPCQ"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGDKLPHRHVFGMGVDDGDNISPHLAMDVPACTK

```


stem_loop
terminator

SPVVCYDADPTGSGMMHVVVNL/PADTRVLPOGSGSLVAMPDGLVOTRIDPEKGTG
YGCAPPKCFETRYIFTYHALDIERIDVDEGASGMVFNHFHSLASASITAMS"

5583. .5605
5583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00013;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggatctgctgttgaagcgacgacg 25
|||||

Db 179 GGATCTGCTGTTGAGCGACGACG 203

RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS

DEFINITION Sequence 6 from Patent EP0798384.

ACCESSION A93679

VERSION A93679.1 GI:6741867

KEYWORDS

SOURCE Escherichia coli.

ORGANISM Escherichia coli

REFERENCE 1 (bases 1 to 5872)

AUTHORS Birch,O., and Brass,J.

TITLE Biotechnological method of producing biotin

JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;

LOCUS AG (CH)

FEATURES

source location/Qualifiers

1. 5872

/organism="Escherichia coli"

/strain="DSM498"

/db_xref="taxon:562"

/clone="PBO30A15-9"

1141. 1156

/standard_name="BIOF RBS"

1154. 2308

/gene="BIOF"

1154. 2308

/gene="BIOF"

1154. 2308

/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"

/EC_number="2.3.1.47"

/codon_start=1

/transl_table=11

/product="DTB SYNTHASE"
/protein_id="CAB69595.1"

5583. .5605
5583. .5644
/db_xref="GI:6741869"

/translation="MSKRRYTGDTVEGTVASCALLAARAGRTAGYKRVASGS
EKTEPEGLRNSDALALQNRSSLSLDYAVVNPYTAEPSPHIISAQGRPLESLVMSAG
LRLEQADWVLYVGAGGWEPLSDPTFADWYTDQLPVILVGVKLCINAMILTA
QVIOHACITLQAGVANDVTPPGKRHAHYMTTLRMIPAPLIGEIPWLAEPENAAAGK
YINLAFVDSITLGFSTRL"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00013;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggatctgctgttgaagcgacgacg 25
|||||

Db 179 GGATCTGCTGTTGAGCGACGACG 203

RESULT 10
A93679 5872 bp DNA PAT 14-FEB-2001
LOCUS

DEFINITION Sequence 1 from patent US 6083712.

ACCESSION A93679

VERSION A93679.1 GI:12812607

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5872)

AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.

TITLE Biotechnological method of producing biotin

JOURNAL Patent: US 6083712-A 1 04-JUL-2000;

LOCUS AG (CH)

FEATURES

source location/Qualifiers

1. 5872

/organism="unknown"

1318 a 1552 c 1695 g 1307 t

ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00013;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggatctgctgttgaagcgacgacg 25
|||||

Db 179 GGATCTGCTGTTGAGCGACGACG 203

RESULT 11

LOCUS A93679 5872 bp DNA PAT 14-FEB-2001

DEFINITION Sequence 6 from patent US 6083712.

ACCESSION A93679

VERSION A93679.1 GI:12812608

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5872)

AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.

TITLE Biotechnological method of producing biotin

JOURNAL Patent: US 6083712-A 6 04-JUL-2000;

LOCUS AG (CH)

FEATURES

source location/Qualifiers

1. 5872

/organism="unknown"

1318 a 1552 c 1695 g 1307 t

ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00013;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggatctgctgttgaagcgacgacg 25
|||||

Db 179 GGATCTGCTGTTGAGCGACGACG 203

RESULT 11

LOCUS A93679 5872 bp DNA PAT 14-FEB-2001

DEFINITION Sequence 6 from patent US 6083712.

ACCESSION A93679

VERSION A93679.1 GI:12812608

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5872)

AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.

TITLE Biotechnological method of producing biotin

JOURNAL Patent: US 6083712-A 6 04-JUL-2000;

LOCUS AG (CH)

FEATURES

source location/Qualifiers

1. 5872

/organism="unknown"

1318 a 1552 c 1695 g 1307 t

ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;


```

VERSION      AE005258.1  GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Escherichia coli O157:H7 EDL933
              Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postel,D.J., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobleck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
              Nature 409 (6819), 529-533 (2001)
TITLE        JOURNAL MEDLINE PUBMED
              11206551
              2 (bases 1 to 13501)
              Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postel,D.J., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobleck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Direct Submission
              Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
              Location/Qualifiers
              1. 13501
              /organism="Escherichia coli O157:H7 EDL933"
              /strain="EDL933"
              /serotype="O157:H7"
              /db_xref="taxon:155864"
              /note="enterohemorrhagic"
              <1..7576
              /note="O-island #36; Region of the EDL933 chromosome not
              homologous to E. coli K-12 MG1655; Cryptic prophage
              CP-933K; includes one copy of the 13 bp direct repeat that
              flanks the prophage"
              66..665
              /gene="lomK"
              /note="20981"
              /gene="lomK"
              /function="putative membrane; Other or unknown (Phage or
              prophage Related)"
              /note="Residues 1 to 199 of 199 are 72.86 pct identical to
              residues 1 to 199 of 199 from GenPept 118 :
              g117532789|gb|AA63231.1|AF151091_2 (AF151091) lom
              [prophage P-Elba]"
              /codon_start=1
              /transl_table=1
              /product="putative outer membrane protein of prophage
              CP-933K"
              /protein_id="AAG55137.1"
              /db_xref="GI:12513752"
              /translation="MKVCAIIISAATCLAVSGVPAMASEHOSTLSAGYLAHTDAPG
              SDDINGINVKRYEFTDITLITSFSTANAEEDKTHYSDTRMHEDIIVRKWTSVMAG
              PSVVENMFSAVAGVAYSRSYFSGDYFRVDNKRRTIDYLDIGSDARYSMTSLAW
              GAGVOENPTESVAADVAYEXSGSDWRTDGFIVGVYK"
              719..2041
              /gene="20982"
              719..2041
              /gene="20982"
              /function="putative structure; Structural component (Phage
              or Prophage Related)"
              /note="Residues 164 to 440 of 440 are 68.79 pct identical
              to residues 381 to 645 of 645 from GenPept 118 :
              g114585436|gb|AAD25464.1|AF125520.59 (AF125520) putative
              tail fiber protein [Bacteriophage 933w]"
              /codon_start=1
              /transl_table=1
              /product="putative tail component of prophage CP-933K"

              /protein_id="AAG55138.1"
              /db_xref="GI:12513753"
              /translation="MGAAVOISGVLRKDGAKPIONCTIOLKARNSTXVNTVASE
              NPDEACRSMVDVGEQYSVTLVNGPFPSPHAGTISYVEDSQPTLNDPLGAMEDAR
              PEALRREDOVNEEARIAHEEKAKKAGAAETESARRAGISAKAKARASANDTSEADSE
              SARQAAESAASAKKSEEAASSSSAQAOKSESISQATDLELSKTAESAGNAADA
              TTTTEKARESAESAQSAEQRIAEDAVNRPVIVGPGPKGEGEPAGDPGGADGE
              RGDTPAGATGRCGPGCDTGTGACGPKGRCGEGETGLTGMAGPGCPKGDTCAGPA
              GPGPKGTGTGACGPGATGPGPKGPGDGGEGTIRRLPMRIETNSYGMFPGDGLI
              TGLFTLPKDAITQVGMFQHLQVRFQGGPQMDVKGDEVSDDGRGE"
              2004..2312
              /gene="20984"
              2004..2312
              /gene="20984"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="Residues 1 to 102 of 102 are 98.03 pct identical to
              residues 28 to 129 of 129 from GenPept 118 :
              g114585437|gb|AAD25465.1|AF125520.60 (AF125520)
              hypothetical protein [Bacteriophage 933w]"
              /codon_start=1
              /transl_table=1
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AAG55139.1"
              /db_xref="GI:12513754"
              /translation="MKAAVIOAELENDMNLRLKIMOSLGGCGKHDCEGNSLTAOLR
              LGPADLESDENGIIPEQDRIYTVLIDADKKQIQCVVRPLQILRADGTWENIGMK"
              2489..3469
              /gene="20985"
              2489..3469
              /gene="20985"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="Residues 13 to 326 of 326 are 55.73 pct identical
              to residues 18 to 331 of 336 from GenPept 118 :
              g116960367|gb|AA63527.1 (AF170176) hypothetical protein
              predicted by Glimmer [Salmonella typhimurium LT2]"
              /codon_start=1
              /transl_table=1
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AAG55140.1"
              /db_xref="GI:12513755"
              /translation="MLSPRTTFHNSVNIIVQSPQVTSFAGKEVELKVIDEKPILF
              OMFEPNPERYKKEDEVIVNTKOHPRVLDVNTNARIESDRMIGIFVGDGFSVNOKTAFS
              KLEDPENMILITREDVDPESMWRKLTSDIYHDIQEOELRTEDKRDDELMLLELR
              ETSKQDSLISMTAKKRNHAMDFERNALALKAGTIFPCTYNTKNGISFEGCITVD
              NMMLITGKIGTIVADGISMHYDRNDSVNIENSAITVNSHHPALLEGLSFPMKSYD
              AHPYVDGLGKGVKVFNEFPLNHNHCFDPIEFNPNITMNTSQTSSW"
              3530..4522
              /gene="20986"
              3530..4522
              /gene="20986"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="No significant matches"
              /transl_table=1
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AAG55141.1"
              /db_xref="GI:12513756"
              /translation="MKIPISLOSNFNSAPAGYSAPFAPRAENAVADYVLDIGKRPL
              SAADLSNVEYSVIRAVHDSRLIDOHVDMGIVLDLSSRQTFRAVSGYAGINERK
              VHIQCTKYRENEYLENSESQKIDDIQSLIHTYNGSSDPSGDSITELGLARVADDELG
              VFSVAPPDIDTSIEMRSWOGILHETIIHHVTCSSDPSGDSITELGLARVADDELG
              WSPVDFKGYAEPEREAHLRLNLMRLQAAHHEHNERAFPERLTISDRYASDPFT
              EYSAVSNIGVFTQOHDFGLAINDLDANOILYHGAPVYETFGDDKHMQ"
              5350..6231
              /gene="20989"
              5350..6231
              /gene="20989"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"

```

```

/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage Cp-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPSYNLCSSNNLRLTLPDNRVLSSRDAAVHSDNGAOV
KGNRTYVATDMKFCYTRKSHSGCFNLHLRGLWPKGELSRKTEVMI.NASPVSAK
ERGIHSHNRDLPVDAAPPLPSVDAVNRSLGPNVAVGNAVYDAADARVIAK
FTTSQSNMEATSEVRCNNOYAGASAKIYNGNDIGIMDKINGESLLNISLPAC
AEHAIDMFRLERQKILFVDTETNVLVYBRAKNEFNPIDISSYVNSDRSSQIMQ
SYHGKODLISVLSKI"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
/function="orf: Other or unknown (Phage or Prophage
related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage Cp-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNVLHSSRSSMDPDIESTLAKASVOMIKNNPOLRF
OGTDNITVOOEALDKIGSTETGRVILNIESIRLSSTVYIHLNSSLGYMAHD
IDAENHRCGSDPFCNNAVEYPCGECISVDPHATIVFHELLHVEHNLGELKVES
SRASQKXSPLLLEARTVGLCASEVLSENKFHEICMPRTSYPDSSALLHDDNT
VLSFQVQRLLPPL"
complement(7668..8144)
/gene="yhbh"
/note="20992"
complement(7668..8144)
/gene="yhbh"
/function="orf: Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
MG1655: B0773
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLKDGLKPLRHVHNGVYDGDNISPLHAMDVPAGTK
SFVATCPDPAIPGSGMMHVVNLPAIDIRVLPQGGSGIVAMPDGLQTRDPFGKG
YDGAAPKKGTHRYITFVHALDVERIDVDGASGAVGVNFHSHIASASTAMPS"
complement(8203..9492)
/gene="b10A"
/note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggaatcgtcttgaacgcagcag 25
|||||
Db 9641 ggaatcgtcttgaacgcagcag 9665

RESULT 15
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
DEFINITION AP002553 BA000007
ACCESSION AP002553.1 GI:13360211
VERSION
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)

AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
TITLE Complete nucleotide sequence of the prophage v71-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
JOURNAL Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE 20198780
REFERENCE 2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
TITLE Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
JOURNAL Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE 20557356
REFERENCE 3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
TITLE Complete nucleotide sequence of the prophage v71-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
JOURNAL Gene 258 (1-2), 127-139 (2000)
MEDLINE 20564182
REFERENCE 4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsuo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shibata,T., Hattori,M. and
Shinagawa,H.
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL DNA Res. 8 (1), 11-22 (2001)
MEDLINE 21156231
REFERENCE 5 (bases 1 to 297816)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp),
URL: http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
COMMENT genome project.
FEATURES
Location/Qualifiers
source 1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
/note="similar to SUCC-ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/translation="MNLHEVQAKQLFAPVGLPAPVGYACTTPREAEASKIGAGPV
VKCOVHAGGRKAGGVVNSKEDIRAFANMLGRRLVYOTDANCOPOVLIYEAT
DIAKEILYGAIVDSSRRVEMASTEGVEIEVAETPHL.IKVALDPLTGPYOG
RELAFKILGKGLVVOOFTKFMGATLFLERDLALTEINPLVYTRKGDILCDKGIKA
DGNALPROPDLREMRDSDOPREAOAOMELNVVALDGNIGCVAGVAGIAGMTDVI
KLHGEFANFLVDGGATKERVIEAKITLISDKVAVALNIFGIVRDLADGIIIG
AAVEGVNPPVAVVLEGNNAELGAKKLADSLNIIAKGLTDAAOOVVAVEGK"

```



```

gene
  /db_xref="taxon:562"
  4..128
  /gene="biob"
  4..9
  /gene="biob"
  20..128
  /gene="biob"
  103..121
  terminator
  /gene="biob"
  /note="early terminator"
  31 a 28 c 34 g 35 t

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 21; DB 1; Length 128;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gagatcgtcttgaagcga 21
Db 24 GGATCTGCTTTGAAGCGCA 44

RESULT 17
AF248314
LOCUS
DEFINITION
  AF248314 8227 bp DNA BCT 24-JAN-2001
  uncultured bacterium pCosAS1.
  uncultured bacterium pCosAS1
  Bacteria: environmental samples.
  1 (bases 1 to 8227)
  Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streitz, W.R.
  Direct Cloning from Enrichment Cultures, a Reliable Strategy for
  Isolation of Complete Operons and Genes from Microbial Consortia
  Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  1133432
  2 (bases 1 to 8227)
  Entcheva, P., Liebl, W. and Streitz, W.R.
  Direct Submission
  Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
  Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
  Location/Qualifiers
    1..8227
    /organism="uncultured bacterium pCosAS1"
    /db_xref="taxon:143793"
    /clone="pCosAS1"
    /note="unknown organism, cosmid clone derived from
    environmental consortium"
    <1..502
    /gene="hutu"
    <1..502
    /gene="hutu"
    /note="putative urocanate hydratase"
    /codon_start=2
    /transl_table=11
    /product="urocanase-like protein"
    /protein_id="AAC53593.1"
    /db_xref="GI:12407618"
    /translation="PRELISFOGIPARICWGLEROKLGLAFNEMVRREVSAPVI
    GRDLDSGVSAPNRETEAMODSDAVSDPMLNLALNTASGATWSLHHGGCGVGF
    SOHAGVIVODGTDAAARLARVLHNDPATGVRRHADGAYDIALIDCAKQGGILPMIA
    AFQGGH"
    502..2031
    /gene="hutu"

```

```

CDS
  502..2031
  /gene="hutu"
  /note="putative"
  /codon_start=1
  /transl_table=11
  /product="histidine ammonia-lyase-like protein"
  /protein_id="AAC53586.1"
  /db_xref="GI:12407611"
  /translation="MNAITLTPGTLTLAQLRQWQOPLQTLDESABAIINDSVACVE
  AIVAEGRVYGINTEGSLIAQRIATHDLENLQRLSVLSHAAGVEPLDDIVRLMV
  LKINSLARGEISGRLSVIOALIALVAGVSYDPAKGSVAGSGLAPLHMSITLIGE
  GKARGLPWLPAATLQKAGLAPVTLAKEGALINGQASTAPRLRLPFAEDLPAS
  AVYCGALTEAVLGSRRPDARIHEVRQKQRTDAALFRVLTDTSLIASHHNCDK
  VDDPYSLRCQPMGACLTQMRQVAEVLIVESNADSDPLVFAENEMVFRGNHAEK
  VMADANLALAIETGALSERRIALMMKQHSQILPPLVRNGVSGFMIAQVTAAL
  ASENKGLCHPHTTSVDKPCRPANQEDHVSNAFPAQRRLMAGNTRGVLAVENTAAC
  QGADLRDGLTSSPLLEQAQSCGEQVAVHDDRFAPDIEAIISELNKGSLVGLLPAP
  L"
  complement(2130..2606)
  /note="ORF1, similar to hypothetical 17.1 kDa protein in
  modC-bioA intergenic region"
  /codon_start=1
  /transl_table=11
  /product="unknown"
  /protein_id="AAC53587.1"
  /db_xref="GI:12407612"
  /translation="WKLISQDLRDGDKLPLNHFVNGMGYEGDNISPHLMDVEPSGK
  SHVAVCYDPAQDPSGMMHWIVANTPADTAVLPQSGSSLVALPEGAETPTDGKAG
  YCGAAPKKEHTRHYFTFVHALVERIEVEDEASGAMGVNHFHTGASITAMYS"
  complement(2651..3940)
  /gene="biob"
  complement(2651..3940)
  /gene="biob"
  /note="7,8-diamino-pelargonic acid aminotransferase"
  /codon_start=1
  /transl_table=11
  /product="DAPA aminotransferase BioA"
  /protein_id="AAC53588.1"
  /db_xref="GI:12407613"
  /translation="WTTDDLAFDRQNIHWPYTSMTQPLPVYVVEANGCELTLASGEO
  LVDGSSWMAALHGYNHPRLNANMAQIDIRSHVAFGGITHPAVUDLCRLVAMTPE
  LECVPADPSVAIVAEVAMKALQVNHAKGESRQETFRNRYHGDFTGAMSCDPDS
  MSLKMGYLPENTLFAFDAPRSDGAMDMWVGLAPLMAARHEIAVLTLEPTYGAG
  GMRVHNPMLKIRRMCDREGILLDAELANGRGRTGLPACBAGATPDLICGKAL
  QVATITMSANTLTTRQVATTISNGEACGCMHPTTMCNPLACAVASESIALISEGMOG
  OGAATESQDLIELAVARDFEVDVRLGALGVETHPVMMALQRFVARGVWRP
  FEKLIYLPMPVYIRPDQLSKLTIRAVVAVOAPAHFTR"
  4029..5069
  /gene="biob"
  4029..5069
  /gene="biob"
  /gene="biob"
  /codon_start=1
  /transl_table=11
  /product="biotin synthase BioB"
  /protein_id="AAC53589.1"
  /db_xref="GI:12407614"
  /translation="MAHDARWTLISQVYALFEKPLLELFFAQQIHROHPPQOITQVST
  LLSITGACPPDCRYCPQSARTKGLLSBRIMEVOULESARQAKNGSTRFCGAMW
  LNPHERDMPYLEQWVGQKAMGLQACMTLGLDHDQARLASGLDYNNHNLDFEPEF
  YGNIITRTYQERLDTLDKVDAGIKVCSGIVGLCEYVDRAGLILQILANLPTPES
  VPINMIVKCTPLADNDVYAFDXIRFXSVAXVMPTXLVRLSTGEONNEOTQAMC
  FMAGANSIFVCGKRLTTPNPREDKDVGLFKRLGINTPOQTAVLTGDNQGHQDLQFLN
  ADTOFYNAATV"
  5066..6223
  /gene="biob"
  5066..6223
  /gene="biob"
  /note="7-keto-8-amino-pelargonic acid synthetase, similar
  to BioB"
  /codon_start=1
  /transl_table=11
  /product="7-kAPA synthetase"
  /protein_id="AAC53590.1"

```



```

/db_xref="GI:12407615"
/translation="MTWQORIDALDARRADALRVTPVENGACRPLIRORRYCNF
SNDYLGISQHEPIIRAWQOQAEYGVSGSGHYSTTHAQLLEELACMLGPPRA
LFIISGFANQALIALTDKEDRIYADRLSHSLSEASHSAQOLRRRANHNDVAQDS
LIDKRGHGOGLVTEGIFSGMDGSDAPLGVIIHAAQROGSMILVDAMHIGVYGPGRG
SAAQSGKELLVTEFGKGVGSAVACSDVAAYLLOFARHLIYSTSMPPAOVAL
LAFNVIRSDENMARQORALIIDOPFAGVRLPPLITASOSAIOTPIVCEANARLHL
AQLRRCGLWVAIRPTVPACTARLRLTLTAHQPAOITLALVEALHYAGE"
6207..6962
/gene="bioc"
/codon_start=1
/db_xref="GI:12407616"
/translation="MSQVNRKQALIAAFGRAGSYSHRDALQROSDGLLAEINTFA
SVLDACGPGSNSIAMRGTGYVTAMDLCAPMLEERRMOADSYLADILEALPLADA
OESLAWSHLAVOXCASTPRALAEIRVYTOLOGKVATITLNGSLPELINOAMQVPTOP
HANRFLSGQOVNDALSGWQRCVAVOTVTLFDALASAMHSLKIGICATHLHAGRAKPL
TRGLOLOLAQAWQOAGKFPPLSYOLFHGIIEBD"
6940..7680
/gene="bioc"
/codon_start=1
/db_xref="GI:12407617"
/translation="MELSWTFRFYTGDTVEYKTVASATLLOARALGRHTAGYKP
VASGSEHTAGLNSDALALQNSLSLSTPANNPTTFAPTSPHIVSDEDEPFEFA
TILSAJRALETQADHWLVESAGGWTPLSOEYTFAMVMERLPLLVGVKLCINH
AMLTQAIRQAGLHFAGVANGVYVPGKRAEYMATLRVLSAPLLGETIPWGEQADT
SAYGOYLDITATPPSAPSSGASGHPAR"
complement(7619..>8227)
/gene="eisa"
/complement(7619..>8227)
/gene="eisa"
/codon_start=1
/product="putative urea/short-chain amide ABC transporter"
/transl_table=1
/product="ABC transporter-like protein"
/protein_id="AA053594.1"
/db_xref="GI:12407619"
/translation="DCXLRTRXGEXNLVSCITAGLPLTASGTWHEAPKAVNAGTEI
XVVPQDRIRFSQITVEENLHIAQMAKGIAGCSLRDVFEEFPELYPLRLQLRASGLND
QOYOLALANLVTRPRLILDEPSRGOLFVYKLAOLILFNPRLGLSVILABOHLIS
LIRVADRFCLVRGVAQGRVSELDPLIAHMCPTASAO"
BASE COUNT 1512 a 2642 c 2605 g 1453 t 15 others
ORIGIN
Query Match 84.0%; Score 21; DB 1; Length 8227;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

REFERENCE 1 (bases 1 to 1041)
AUTHORS Wu, C.H., Chen, H.Y. and Shiu, D.
TITLE Isolation and characterization of the Erwinia herbicola bio operon
and the sequences of the bioA and bioB genes
JOURNAL Gene 174 (2), 251-258 (1996)
MEDLINE 97045821
REFERENCE 2 (bases 1 to 1041)
AUTHORS Shiu, D.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1995) David Shiu, Biology, Natl. Sun Yat-sen
University, Kaohsiung, Taiwan, ROC
FEATURES
Source 1..1041
/organism="Pantoea agglomerans"
/strain="Eno10"
/db_xref="taxon:549"
1..1041
/gene="biob"
1..1041
/gene="biob"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AAC44534.1"
/db_xref="GI:1228114"
/translation="MAHHAHMTLSQVETLFDKPLDLLFEAOTTHROHPDPROVST
LSTITGACPEDCKYCPQSSRYKGTGLFERLMEYQVDSARKKNGSRFCGAM
KNPHERDMPYLEQWQVYKAMGLACMTGLTDETOARLASADLYNNHLDPSPE
VGNITTRTYOERLDTLDKRVDAIGKVGSGIVGIGETVKRAGLLQALNLPDPS
VPINMLVRYGTPLADNDVDAFDRTIYAVARIMPTSPVRLSAGEBOMEQOAMC
PMAGANSIEYCGKILTTTPNEBDKQVQLFRKLGNPQOTDVMTDNEQOKLEBOIFN
ADTDFYNAAL"
BASE COUNT 230 a 322 c 308 g 181 t
ORIGIN
Query Match 72.0%; Score 18; DB 1; Length 1041;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 19
LOCUS E00893 1121 bp DNA PAT 29-SEP-1997
DEFINITION Genomic DNA encoding biotin Synthetase.
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono, Y., Kojima, T. and Kimura, H.
TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE. BACTERIUM CONTAINING SAME AND
PRODUCTION OF BIOTIN
JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
NIPPON SODA CO LTD
OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PT HIRONO YOSHITAKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
C12R1:19);
CC strandedness: Double;
CC topology: Linear;

```

CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli N6101;
 CC Feature is identified by experimental;
 FH Key Location/Qualifiers
 FT CDS 42..1079
 /product="biotin synthetase".
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 289 a 296 c 325 g 211 t
 ORIGIN

Query Match 72.0%; Score 18; DB 6; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtctgtgaagc 18
 |||||||
 Db 104 GCATCTGCTGTTGAAGC 121

RESULT 20
 AF250347/c 1044 bp mRNA VRT 01-JUL-2001
 LOCUS Xenopus laevis transcription factor NKX-2.1 mRNA, complete cds.
 DEFINITION
 AF250347
 ACCESSION
 AF250347.1 GI:14578835
 KEYWORDS
 SOURCE African clawed frog.
 ORGANISM
 Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopodidae; Xenopus.
 1 (bases 1 to 1044)
 Hollemann, T. and Pieler, T.
 Xkx-2.1: a homeobox gene expressed during early forebrain, lung
 and thyroid development in Xenopus laevis
 Dev. Genes Evol. 210 (11), 579-581 (2000)
 MEDLINE 2110556
 JOURNAL 1180810
 2 (bases 1 to 1044)
 Hollemann, T. and Pieler, T.
 Direct Submission
 Submitted (29-MAR-2000) Entwicklungsbiologie, University of
 Goettingen, Humboldtallee 23, Goettingen, Lower Saxony 37073,
 Germany
 FEATURES
 source Location/Qualifiers
 1..1044
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 1..29
 1..1044
 /note="Xkx-2.1"
 /codon_start=1
 /product="transcription factor NKX-2.1"
 /protein_id="AAK69025.1"
 /db_xref="GI:14578836"
 /translation="MSMSPKHTTPPSVDILSPLESYKRVAMEGAGLPLTAAYR
 OSOVSQSMOQHMGHNGPVSAAVHMTAGVQLSHITMGVCGNGLGNGMSLPLP
 YODTMRNSATGMYGANPDPRFTISRFGPSGGMNGLGNGMSLPLP
 TPRRRVLFSSQAVYELERFRKOKYLSAPREHLASHTITPTQVKTWPNHKKM
 KROAKDKASQOQODNSCCOQOSRRRAVVLVKGDKPCOAGSNTPPAALQSHQOO
 TATATITTTNGICPHOSHOTSAGOSPDLVPHSPSSSLQNOVYLSLHNSSSDYGS
 AMSSTLILXRTW"
 BASE COUNT 277 a 311 c 265 g 190 t 1 others
 ORIGIN
 Query Match 68.0%; Score 17; DB 5; Length 1044;

Best Local Similarity 100.0%; Pred. No. 8.6;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaagcgc 20
 |||||||
 Db 556 TCTGCTGTTGAAGCGC 540

RESULT 21
 AF281080/c 1614 bp mRNA VRT 01-OCT-2000
 LOCUS Xenopus laevis homeobox transcription factor NKX2-1 (NKX2-1) mRNA,
 complete cds.
 DEFINITION
 AF281080
 ACCESSION
 AF281080.1 GI:10442649
 KEYWORDS
 SOURCE African clawed frog.
 ORGANISM
 Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopodidae; Xenopus.
 1 (bases 1 to 1614)
 Small, E.M., Vokes, S.A., Garriock, R.J., Li, D. and Krieg, P.A.
 Developmental expression of the xenopus Nkx2-1 and Nkx2-4 genes
 Mech. Dev. 96 (2), 259-262 (2000)
 MEDLINE 20417824
 JOURNAL 20417824
 2 (bases 1 to 1614)
 Small, E.M., Vokes, S.A. and Krieg, P.A.
 Direct Submission
 Submitted (21-JUN-2000) Cell Biology and Anatomy, University of
 Arizona Health Sciences Center, 1501 N. Campbell Ave., Tucson, AZ
 85724, USA
 FEATURES
 source Location/Qualifiers
 1..1614
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /note="expressed in ventral forebrain, lungbud primordia
 and thyroid anlage"
 1..1614
 /gene="NKX2-1"
 162..1205
 /gene="NKX2-1"
 /note="Thyroid transcription factor; TTF-1; T/ebp; chicken
 and mouse Nkx2-1 ortholog; contains homeodomain"
 /codon_start=1
 /product="homeobox transcription factor Nkx2-1"
 /protein_id="AAG17405.1"
 /db_xref="GI:10442650"
 /translation="MSMSPKHTTPPSVDILSPLESYKRVAMEGAGLPLTAAYR
 OSOVSQSMOQHMGHNGPVSAAVHMTAGVQLSHITMGVCGNGLGNGMSLPLP
 YODTMRNSATGMYGANPDPRFTISRFGPSGGMNGLGNGMSLPLP
 TPRRRVLFSSQAVYELERFRKOKYLSAPREHLASHTITPTQVKTWPNHKKM
 KROAKDKASQOQODNSCCOQOSRRRAVVLVKGDKPCOAGSNTPPAALQSHQOO
 TATATITTTNGICPHOSHOTSAGOSPDLVPHSPSSSLQNOVYLSLHNSSSDYGS
 AMSSTLILXRTW"
 BASE COUNT 475 a 407 c 399 g 333 t
 ORIGIN

Query Match 68.0%; Score 17; DB 5; Length 1614;
 Best Local Similarity 100.0%; Pred. No. 8.7;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaagcgc 20
 |||||||
 Db 717 TCTGCTGTTGAAGCGC 701

RESULT 22
 AE001220 14268 bp DNA BCT 16-JUL-1998
 LOCUS Treponema pallidum section 36 of 87 of the complete genome.
 DEFINITION
 AE001220 AE000520
 ACCESSION

```

VERSION      AE001220.1  GI:3322705
KEYWORDS     Treponema pallidum.
SOURCE       Treponema pallidum.
ORGANISM     Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
REFERENCE   1 (bases 1 to 14268)
AUTHORS     Fraser,C.M., Norris,S.J., Weinstein,G.M., White,O., Sutton,G.G.,
            Dodson,R., Gwin,M., Hickey,E.K., Clayton,R., Ketchum,K.A.,
            Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J.,
            Khalak,H., Richardson,D., Howell,J.K., Chidambaram,M.,
            Utterback,T., McDonald,L., Artlich,P., Bowman,C., Cotton,M.D.,
            Venter,J.C. et al.
            Complete genome sequence of Treponema pallidum, the syphilis
            spirochete
            Science 281 (5375), 375-388 (1998)
JOURNAL      98332770
MEDLINE      2 (bases 1 to 14268)
REFERENCE    Fraser,C.M., Norris,S.J., Weinstein,G.M., White,O., Sutton,G.G.,
            Dodson,R., Gwin,M., Hickey,E.K., Clayton,R., Ketchum,K.A.,
            Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J.,
            Khalak,H., Richardson,D., Howell,J.K., Chidambaram,M.,
            Utterback,T., McDonald,L., Artlich,P., Bowman,C., Cotton,M.D.,
            Venter,J.C., Smith,H.O. and Venter,J.C.
            Direct Submission
            Submitted (06-MAR-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source       1. 14268
            /organism="Treponema pallidum"
            /db_xref="taxon:160"
            /gene="TP0421"
            /note="similar to GB:AE000783 percent identity: 27.26;
            identified by sequence similarity: putative"
            /transl_table=1
            /product="conserved hypothetical protein"
            /protein_id="AAC65409.1"
            /db_xref="GI:3322706"
            /translation="MRGWCILSRSSVAGAGALLLFEVAGNOEHAPPVVDALAGEF
            RGVAVYRCAGFNDALDLEFALASAPRNPDLLEMLGNVYRSIGIEGALHOGARD
            LGYGALLRNKIEVVOORHPDAPSDALHFSSESFAVRRTVLEFRPLSLCALAD
            GTFWMASRGSNELLRDPVNGRVYARTRGPEVQEDREDFVLTQSSGLVSEFASDPTC
            RLTRGKFLKSPGKGRVGOQLIGPOLATDRIDTYVDGNAVAVRAPDQAPLPH
            FGOKSARFSGFSAAGIATIDELVYADALKGATIDYTAGNVRTLVAEGLKHVES
            VRAMNGRLVSLPNEVAVDVGLASLHTTARLGNAPIRLTAVPDANGVYLADYKNE
            RIETVRSISELACGLFVHFERVHADRPSTVAVSVTRGTPVGLDVSNEFLTEEH
            RPYSEMRLQGAAYLINDTCSLITVDRSPATEOEMRLVTRVIGLQAVGSGGTLVSYS
            AAOSEVLEGTFTTAQAOAPLRAKASSSOMREDICRLAAGTLVADKRAVAVFSL
            GEIQTDFQAOYGTDLAAVLTINNETSVYIQTORTVLPVELGLTVOKTRCAVYSVYAL
            EGCGLAVAHVSGAAGTYALYSRSLPTDFGRAVMEAEVRVLSRSGRDEGTGFAPL
            E"
            2133..3158
            /gene="TP0422"
            2133..3158
            /gene="TP0422"
            /note="hypothetical protein: identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=1
            /product="T. pallidum predicted coding region TP0422"
            /protein_id="AAC65418.1"
            /db_xref="GI:3322715"
            /translation="MRLSDRLQMGKMSAVVRLCVLSFQVGLAEPALAGW
            ENATRIVAIPTADAMFSLMLVHKPYAVYVEVAVFLSLSLSDARADVLPAPAK
            AERVDIVSFGVRVQPLVLCRWGDALFCSFORVYGRSGSVSEOLPVPRTDPL
            FGVWIEGSAEALPAYSAPGELNAPFLPEANVRRVWRDALFAADRAVFTDPL
            GVTTEIPIYORGAAYVSCITGSRVYRNPTOTGTVEYRTSSDGRMLRRRAGDS
            YAVGAVYPROGFLDESALALRVGDFPLATIGPEPLKSAVHDVTAETDAHARRARAPRA
            PLVVPDF"
            gene
            CDS
            3183..3989
            /gene="TP0423"
            3183..3989
            /gene="TP0423"
            /note="similar to GB:AE000783 percent identity: 39.66;
            identified by sequence similarity: putative"
            /transl_table=1
            /codon_start=1
            /product="conserved hypothetical protein"
            /protein_id="AAC65410.1"
            /db_xref="GI:3322707"
            /translation="MRVFSVMKQDKGEREAGALQACIGTPERHOFALVGRSGS
            GKSFRADIVMRDEIAFVIDGILIREDKTIAGRSKQERTLLAALRVALLFETMRR
            VVARVLYFLGSONKRVLLIGTSEMRVRIALRVLEPGHITRIEDIASSEIACA
            LTRHVEGTHYTPVSGEVRKSYKPIYERIKLLREAGAEIRGRANAHIMHEGLKR
            ACSGAHMHPEKSLVRRPFCNLRVAVATGTDASVYVVPHPHELALQO"
            4192..4890
            /gene="TP0424"
            4192..4890
            /gene="TP0424"
            /note="similar to GP:151015 percent identity: 37.59;
            identified by sequence similarity: putative"
            /transl_table=1
            /codon_start=1
            /product="V-type ATPase, subunit E, putative"
            /protein_id="AAC65411.1"
            /db_xref="GI:3322708"
            /translation="MLGSRREARIVRAAREARIVRAAREARIESSIALAISO
            ASRVNLSFQDSVTRSLRATISMETAQAVAGVIRLELIPVYSAMVQABEDKLEILS
            PADLTLEGFCAALQEBOLSAVGLKRSDDCLTRACFRIVPAEGSYTFPSAAVVAQLES
            SVSARVAEVLSTLRSCDFLLPYNNATISSMRKASFPALSGFGLCATFPLGASAC
            CGAMKYLIVSRAGA"
            4868..5128
            /gene="TP0425"
            4868..5128
            /gene="TP0425"
            /note="hypothetical protein: identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=1
            /product="T. pallidum predicted coding region TP0425"
            /protein_id="AAC65419.1"
            /db_xref="GI:3322716"
            /translation="MCPVADSTGSRVLDKMGFEYALRCALAREADRLMDTPYE
            VROCVAQESVARRAVAVSLDDPLEAGVFPERALFLRCGV"
            5270..7039
            /gene="TP0426"
            5270..7039
            /gene="TP0426"
            /note="similar to GB:AE000783 percent identity: 63.40;
            identified by sequence similarity: putative"
            /transl_table=1
            /codon_start=1
            /product="V-type ATPase, subunit A (atpA-1)"
            /protein_id="AAC65412.1"
            /db_xref="GI:3322709"
            /translation="MTQTKGIYSAVNGMVSVEPGEVYSLNEGVYVHGNARLAEIIL
            RYGRADLOVPEITRGVSGDVGEFTGDLISVLEGPDLQGVYQDGNPLPLAEKV
            GYFLRGVYLPALSRSTSEMFTPHYSVEERVGRVGLVETPEGALKHIMVFNHGDG
            YEVVLIQTACTRYNDVITARYVDAGQHEHELTMARMPVKKRPVHCYARLRKPTPEIYV
            STRITDFEPVAKGGCTYCIPGPFAGKTVLQHSTRNADVYVYTAACGERAGEVET
            LREFPDLDPRTGRSLMERTVIVCMTSSMPVASVYTGVTAEVYRQMGDLVLL
            ADSTSRMAOLREMSGRLEIIPGEAFPAYLESCIAAYERAGVRLRSGKGSVYTG
            GVSPAGKNEEPVQATLKVYGAFHGLSRESDARVPAVPAVPLDSMSKYPYSDAPRA
            VAYGSLRKGAEVQOMRMVVGEGTSMEDFLVYTKGSLDSVYLOQNSKPYVDVAVP
            VAROKHCVATVWRVIGSVIAPSKDARATYPSKLGHMTIDYNCAMNEAFVEKKEI
            RALQGESKTKIDSEADGILKME"
            7043..8335
            /gene="TP0427"
            7043..8335
            /gene="TP0427"
            /note="similar to GB:AE000783 percent identity: 72.96;
            identified by sequence similarity: putative"

```

Query Match	Best Local Similarity	Matches	17: Conservative	0: Mismatches	0: Indels	0: Gaps
Qy 7 gctgtttgaagcagc 23 	68.0%; Score 17; DB 1; Length 1426;	100.0%; Pred. No. 9.3;				
Db 8188 GCTGTTTGAAGCGCAGC 8204						
RESULT 23 AX120335/C LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM	AX120335 Sequence 251 from Patent EP1108790. AX120335 AX120335.1 GI:14037050 Corynebacterium glutamicum. Corynebacterium glutamicum Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae; Actinomycetales; Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.	801 bp DNA PAT 11-MAY-2001				

FEATURES	source	location/Qualifiers
AUTHORS	(bases 1 to 801)	
TITLE	Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K., Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.	
JOURNAL	Novel polynucleotides Patent: EP 1108790-A 251 20-JUN-2001; KYOWA HAKKO KOGYO CO., LTD. (Jp)	
FEATURES	1. 801	
BASE COUNT	184 a 221 c 215 g 181 t	
ORIGIN	1. 801	
Query Match	64.0%: Score 16; DB 6; Length 801;	
Best local Similarity	100.0%: Pred. No. 34;	
Matches	16; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Qy	8 ctgtttgaagcgacg 23	
Db	784 CTGTTGAAGCGCAGC 769	
RESULT	24	
LOCUS	AX064067 924 bp DNA PAT 24-JAN-2001	
DEFINITION	Sequence 349 from Patent WO0100843.	
ACCESSION	AX064067	
VERSION	AX064067.1 GI:12541779	
KEYWORDS	Corynebacterium glutamicum.	
SOURCE	Corynebacterium glutamicum.	
ORGANISM	Corynebacterium glutamicum	
REFERENCE	1 (bases 1 to 924)	
AUTHORS	Pompejus, M., Kroege, B., Schroeder, H., Zelder, O. and Haberhauer, G.	
TITLE	corynebacterium glutamicum genes encoding metabolic pathway proteins	
JOURNAL	Patent: WO 0100843-A 349 04-JAN-2001;	
FEATURES	BASE AKTENGSELISCHAFT (DE)	
source	Location/Qualifiers	
CDS	1. 924	
	/organism="Corynebacterium glutamicum"	
	/db_xref="taxon:1718"	
	101. 904	
	/note="RXA01757"	
	/codon_start=1	
	/transl_table=1	
	/protein_id="CAC25274.1"	
	/db_xref="GI:12541780"	
	/translation="MPTASPIYDVYVVGATISGLIAFOLLDRAGLNTKCFEACSRVGGRAVVSQDSDFLDIGATWFMENPELVQOLVNNLIGTFPOAIEDDALFETLVDPASRRIRGNFIDASGRFOGASSLALGLAOLKPGVLEIGDPVHSISEDGELIVSSKQIVRAAKHIVPAPALAEELIGFTLIDLPADVKKANPHQIAYMMNAKSKYTLPTQASAGSGHLEFQOPLGCRHMASTEVATFGGGLLEAVRAGIQALQGTGFNLKS"	
BASE COUNT	214 a 252 c 239 g 219 t	
ORIGIN	214 a	
Query Match	64.0%: Score 16; DB 6; Length 924;	
Best local Similarity	100.0%: Pred. No. 34;	
Matches	16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	8 ctgtttgaagcgacg 23	
Db	884 CTGTTGAAGCGCAGC 869	
RESULT	25	
LOCUS	YL1012084 6887 bp DNA PLN 23-SEP-1999	
DEFINITION	Yarrowia lipolytica pex10 gene for peroxin 10.	

ACCESSION AJ012084
 VERSION AJ012084.1 GI:5921565
 KEYWORDS peroxin 10; pex10 gene.
 SOURCE Yarrowia lipolytica.
 ORGANISM Yarrowia lipolytica
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetaceae; Dipodasacaceae; Yarrowia.
 REFERENCE 1 (bases 1 to 6887)
 AUTHORS Le Dall, M.T., Gallard, C. and Nicaud, J.M.J.
 TITLE PEX10 is essential for peroxisome biogenesis in Yarrowia lipolytica
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 6887)
 AUTHORS Nicaud, J.M.
 TITLE Direct Submission
 JOURNAL Submitted (20-SEP-1999) Nicaud J.M., Laboratoire de Genetique des
 Microorganismes, INRA-CNRS, BP 01, Tallevrat-Grignon, F78850, FRANCE
 FEATURES
 source
 1..6887
 /organism="Yarrowia lipolytica"
 /strain="W29"
 /db_xref="taxon:4952"
 /complement(2293..3357)
 /gene="pex10"
 /complement(2293..3357)
 /gene="pex10"
 /function="essential for peroxisome biogenesis"
 /codon_start=1
 /evidence=experimental
 /product="peroxin 10"
 /protein_id="CA56484.1"
 /db_xref="GI:5921566"
 /translation="MSDNTIKKPIRPPIRTERLPYAGAEIIRANOKDHFESVLE
 OHLVITLQMKKVRREHLHOKYKELLETASKPAGVIGCTLVSKGSEYTNIMATIDRER
 ALPGVRRRGYVLTLPYLFVRWTKLRKLMREYPHLYVDEDEPPSPETPKER
 VKTVEVNRKFTALEGFTAIHLAIFVYGVYQSLRWKRVYVGHRLDKNEPRIG
 YMLGLIFARFATISVOTGREYLALEKSVKAGEKDEKAVVPPKKSIPFIE
 DREGEDKIDLEDPOLKFIPEASRACLTCLSTISAPACPGHFCGDCISEWVRE
 KRCPLRCGVGEONLPIR"
 BASE COUNT 1793 a 1841 c 1485 g 1768 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 8; Length 6887;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 gatctgctgttgaag 17
 ||||||||||||
 Db 4051 GATCTGCTGTTGAG 4066
 RESULT 26
 AE002054 11823 bp DNA BCT 22-NOV-1999
 LOCUS Deinococcus radiodurans R1 section 191 of 229 of the complete
 DEFINITION chromosome 1.
 ACCESSION AE002054 AE0009513
 VERSION AE002054.1 GI:6460010
 KEYWORDS
 SOURCE Deinococcus radiodurans.
 ORGANISM Deinococcus radiodurans
 Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 REFERENCE 1 (bases 1 to 11823)
 AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
 Dodson, R.J., Haft, D.H., Gwin, M.L., Nelson, W.C., Richardson, D.L.,
 Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
 Vamathevan, J.J., Lam, P., McDonald, L., Uitterback, T., Zalewski, C.,
 Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M., et.al.
 TITLE Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1
 JOURNAL Science 286 (5444), 1571-1577 (1999)
 MEDLINE 20036896
 REFERENCE 2 (bases 1 to 11823)

AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
 Dodson, R.J., Haft, D.H., Gwin, M.L., Nelson, W.C., Richardson, D.L.,
 Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
 Vamathevan, J.J., Lam, P., McDonald, L., Uitterback, T., Zalewski, C.,
 Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
 Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
 Smith, H.O., Venter, J.C. and Fraser, C.M.
 TITLE Direct Submission
 JOURNAL Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source
 1..11823
 /organism="Deinococcus radiodurans"
 /strain="R1"
 /db_xref="taxon:1299"
 /chromosome="1"
 103..390
 /gene="DR2207"
 103..390
 /gene="DR2207"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AA11762.1"
 /db_xref="GI:6460017"
 /translation="MDMSHPFEARALOLAYEMTAHNTYEGKGGGCHNEARADIRRL
 FDICGHDPEDPAHSHKAGRKRAEAKRRKKFADLPVAEALGNS"
 complement(449..829)
 /gene="DR2208"
 complement(449..829)
 /gene="DR2208"
 /note="similar to SP:060003 percent identity: 48.36;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="lactoylglutathione lyase, putative"
 /protein_id="AA11756.1"
 /db_xref="GI:6460011"
 /translation="MRALETGLYADDLISAAEFYSGVGLTLHSKVAGRHLEFRDGS
 MLIFNPASAKKPGDVPDSHOCKPGCHACIAIPROGTAWMETRLRGHLDVTRAMCGR
 GESLTFDDPGNVLELAPASTWGL"
 complement(993..1904)
 /gene="DR2209"
 complement(993..1904)
 /gene="DR2209"
 /note="similar to PID:2338556 percent identity: 50.20;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="Subd-related protein"
 /protein_id="AA11757.1"
 /db_xref="GI:6460012"
 /translation="MSAHTDMDGNDGSDHEDMDADPCQKLRKTRKPOGRKKA
 DLTGERTDIADDTIRLLDGLTEVVCETSKSEKNTAVTAETVIRPALIKTIRDA
 ARSFKNDAIYREGGVLDORARAMESKRSKRLDMLQFDWGAETVAHLMDQAGLSV
 PEPLVGHKYTAETPAVLMRLIGTAESPAPRLSDVLTPEEARRMOOSVOCMDL
 LRLGYAHGDSYTNLWMENTVTITIDFOLTRRANPNKELLRRDAOSIATSFRRHCL
 DETAGCTLRVQRRLGAPPRVLP"
 2390..2773
 /gene="DR2210"
 2390..2773
 /gene="DR2210"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AA11763.1"
 /db_xref="GI:6460018"
 /translation="MKGYCSGLLLALLALSPLOSRAQOATSCNYPVGTPODYCVNK
 VTLQADTEMTAYOKLRLKLPYSAOQTLRGTORAMMAKRDNDVDGYTVVYVNCYHDMT
 VSRINFLNDRLEPCNAAGCPSKLR"
 2813..3505

gene

REFERENCE
AUTHORS

2 (bases 1 to 13856)
 Simpson,A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,
 Alvares, R., Alves, L.M.C., Araya, J.E., Baiz, G.S., Baptista, C.S.,
 Barros, M.R., Bonaccorsi, E.D., Bordin, S., Bove, U.M., Brites, M.R.S.,
 Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M.,
 Carrier, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R.,
 Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E.,
 Docena, C., El-Dorry, H., Facincani, A.P., Ferreira, A.J.S.,
 Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C.,
 Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S.,
 Gomes, S.L., Gruber, A., Ho, P.L., Hohelsel, J.D., Junqueira, M.L.,
 Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F.,
 Lambis, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A.,
 Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N.,
 Madeira, H.M.F., Martino, C.L., Marques, M.V., Martins, E.A.L.,
 Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C.,
 Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A.,
 Nascimento, A.L.T.O., Netto, L.E.S., Nhani, J.R., Nobrega, F.G.,
 Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C.,
 Palmeri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Rodrigues, V.,
 Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodriques, V.,
 de M. Rosa, A.J., de Rosa, Jr., V.E., de Sa, R.G., Sanceli, R.V.,
 Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva
 Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de
 Souza, A.A., de Souza, A.P., Terezzi, M.F., Truffi, D., Tsai, S.M.,
 Tsubako, M.H., Vallada, H., Van Sluys, M.A., Verjovsky-Almeida, S.,
 Vettore, A.L., Zago, M.A., Zatz, M., Melanlis, J. and Setubal, J.C.
 Direct Submission
 Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
 Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
 13083-970, Brazil

FEATURES
source

location/Qualifiers
 1..13856
 /organism="Xylella fastidiosa 9a5c"
 /db_xref="taxon:160492"
 /clone="9a5c"
 /complement(195..803)
 /gene="XF2101"
 /complement(195..803)
 /gene="XF2101"
 /note="similar to GI13552989 (percent identity: 30 %/query
 alignment coverage: 87.6 %/subject alignment coverage:
 67.6 %): identified by sequence similarity: putative; ORF
 located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF84900.1"
 /db_xref="GI:9107229"
 /translation="MSRCITGYRMWQWICVAVLVFPMVAAGMIWIGORDQAAAS
 DVLIVGAAYDAKPSVPFPERIIVYOHGAPLITFGCGSSAFAASOVAR
 RYALRYVPRALIIETVSTRNOLIOAOILMVYGLARRAIVSDPLIMARRALRLCR
 ELIDILASATPSTFRPSRLRFLDVEYFFRRDLIVIAN"
 /complement(800..1042)
 /gene="XF2102"
 /complement(800..1042)
 /gene="XF2102"
 /note="hypothetical protein: identified by sequence
 similarity: putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF84901.1"
 /db_xref="GI:9107230"
 /translation="MARVIMYSSFAISVSKICGLVTGMLLHAIVFGGIGCFRAL
 CCDAVWLCHVDISARLSGCGSAGFFICGVSIVLYVS"
 1397..1732
 /gene="XF2103"
 1397..1732
 /gene="XF2103"
 /note="hypothetical protein: identified by sequence
 similarity: putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1

gene
CDSgene
CDSgene
CDS

1397..1732
 /gene="XF2103"
 1397..1732
 /gene="XF2103"
 /note="hypothetical protein: identified by sequence
 similarity: putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1

gene

CDS

gene
CDSgene
CDSgene
CDS

tRNA

/transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF84902.1"
 /db_xref="GI:9107231"
 /translation="MKARLRLPLILMSAISGCAATYNSNGYHGTPVEVHYPPNY
 PHGGMDLVEVPLIPPLYPRDFYKTHILHPWQRPHDPSRPLPRKPIKQEPNHE
 MGLDKDKPI"
 1914..2024
 /gene="XF2104"
 1914..2024
 /gene="XF2104"
 /note="hypothetical protein: identified by sequence
 similarity: putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF84903.1"
 /db_xref="GI:9107232"
 /translation="MNADDDMLCSELEIMKICPCVDTLKHQLAAR"
 /complement(2067..2225)
 /gene="XF2105"
 /complement(2067..2225)
 /gene="XF2105"
 /note="hypothetical protein: identified by sequence
 similarity: putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF84904.1"
 /db_xref="GI:9107233"
 /translation="MAASREBAMEFYFCVIFRNALFRVGTYYADVAPAAALRL
 QAORSGLE"
 2271..4160
 /gene="XF2106"
 2271..4160
 /gene="XF2106"
 /note="similar to SPIP25756 (percent identity: 65 %/query
 alignment coverage: 98.6 %/subject alignment coverage:
 98.4 %): identified by sequence similarity: putative; ORF
 located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="glucose inhibited division protein A"
 /protein_id="AAF84905.1"
 /db_xref="GI:9107234"
 /translation="MTDSFYGYDVIVIGAGHNGSEALAAARTGAHTLLTHNIETIG
 AMSCNPAIGIGKHLVKEIDALGAMHAADAAGIOWRTLNASKPAPVATRCADR
 ALYRAIRIOMITIGQARLNTFOAEVDLLFEGDTVGATGTHGLHKKAVAVLTAGTFL
 AGRTIHGPIOTYAAGRMGDDPATMLAARLERLFYARLRKTGPPRIIDRTLTNYMOE
 QPQDARPPMSTIGNSASHPPPOVSCWITOTTERTEIRIIRALHRSPLXSGOIBGPR
 YCPSEIDKVRFAEKNSHOIEVEPGLVNIDLYPNGISTSLPEVDLEVNSIRGEFO
 AHTRPGVYIEVDFEPRGLKASLETKALAGLEFPAGONGTGYEEAAGGLALNA
 AHRYRGVLSMPRPDOAYLGVLYVDLLITPGTNEPVRMFTSRAEYELQJREDAADARL
 ATCRDLGILIDDAARHAFNAKQAVARECEKSAFATGNAAGREVKETLGVLSRET
 NIIDLRKPELDIYALIMRPSYKSGVDAOVNAEOVEISVYKAGYINROSEETRCORH
 ENALPLEFDYAVAVRGSTEVYLOKLIQHIDPOTVGOAQRIPGHTPAISILLVHLERM
 KMRVA"
 /complement(4348..4560)
 /gene="XF2107"
 /complement(4348..4560)
 /gene="XF2107"
 /note="hypothetical protein: identified by sequence
 similarity: putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF84906.1"
 /db_xref="GI:9107235"
 /translation="MPLLMCYHTHPLIDELVEAMRAGVIRSMYNAIKODSSGITE
 RLVPKWFWSGRASKIDATAAILASV"
 4928..5001
 /gene="tRNA-Cys"
 /note="found by tRNAscan"

```

gene      /product="tRNA-Cys"
           4928..5001
gene      /gene="tRNA-Cys"
           5207..5392
           /gene="XF2108"
           5207..5392
CDS       /gene="XF2108"
           /note="hypothetical protein; identified by sequence
           similarity; putative; ORF located using Glimmer/RBSfinder"
           /codon_start=1
           /transl_table=1
           /product="hypothetical protein"
           /protein_id="AAF84907.1"
           /db_xref="GI:9107235"
           /translation="MQADGTTITGLDSTKRWATAFLSGONPLSIDKSPFVLGAALIK
           NSAKNDTTRAPAKTA"
gene      complement(5566..5892)
           /gene="XF2109"
           complement(5566..5892)
           /gene="XF2109"
           /note="hypothetical protein; identified by sequence
           similarity; putative; ORF located using Glimmer/RBSfinder"
           /codon_start=1
           /transl_table=1
           /product="hypothetical protein"
           /protein_id="AAF84908.1"
           /db_xref="GI:9107237"

Query Match      64.08; Score 16; DB 1; Length 13856;
Best Local Similarity 100.08; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 gacgcgcgtttgaag 17
Db      2652 GATCGTGTGTTTGAAG 2667

RESULT 28
CBRG01D9/c      CBRG01D9      40031 bp      DNA      INV      02-MAY-1996
LOCUS      Caenorhabditis briggsae cosmid G01D9.
DEFINITION
ACCESSION      U56248
VERSION      U56248.1
KEYWORDS      GI:1293789
SOURCE      Caenorhabditis briggsae strain-Gujarat G16.
ORGANISM      Caenorhabditis briggsae
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Feloderinae; Caenorhabditis.
Wu, X. and Le, T.T.
The sequence of C. briggsae cosmid G01D9
unpublished (1996)
REFERENCE      2 (bases 1 to 40031)
JOURNAL      Washington University Genome Sequencing Center.
AUTHORS      The C. briggsae Genome Sequencing Project
TITLE      unpublished (1996)
JOURNAL      3 (bases 1 to 40031)
REFERENCE      Waterston, R.
AUTHORS      Direct Submission
TITLE      Submitted (23-APR-1996) Robert Waterston
COMMENT      Genome Sequencing Center
      Department of Genetics, Washington University,
      St. Louis, MO 63110, USA
      e-mail: mmarrat@wustl.edu
      NEIGHBORING COSMID INFORMATION:
      The neighboring cosmids, chromosome and orientation of G01D9 are
      unknown.
NOTES:
Coding sequences below are predicted from computer analysis, using
the program GeneFINDER (P. Green and L. Hillier, ms in preparation).
Location/Qualifiers

```

```

source
1..40031
/organism="Caenorhabditis briggsae"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="G01D9"
10..14844
/gene="G01D9.5"
join(10..816,859..1074,1134..2667,2729..2930,2975..13571,
13615..13860,13905..14078,14121..14306,14353..14400,
14449..14558,14609..14693,14746..14844)
/gene="G01D9.5"
/note="Similar to polyketide synthase."
/codon_start=1
/evidence="not_experimental"
/protein_id="AAF84909.1"
/db_xref="GI:1293790"
/translation="MDSLTDFVFNRLNKKYFPEDEITASGVDFPYVDELAADHVRK
KVSFPLPRASGEYRWGVBNP LONEPEVELOSTYTKNSPTGRKLDYLWLSGLQ
LSNPSSNFTLALNDIGDQDELRMLAEKQNLINPKLDSSSTETLWLSLNFVT
LSKSKIKCFNFGSEFLGNSIARAFMKTISEKYLPLSFMDKIQOVSFTDSSSI
PQNWLTGSLGIGFEICKPLNGSAENVILISROPADVYAEIKTWTAVHTIAD
VNNKKRLLELTRVAVDVGTGLHSAGVLKDSKIEROTDSFNEVFKPQACQWLEI
EKHSYKIEHFIMSSFTACGNGEQLWGVANAFLEHNRORRRECKSGAIDWGM
IDTGMATNPVAKFLADIGLQHNKDKLTKACITKPTETVAVANIAMVITLKNK
DIPROLIINGILPESTVITENGNDSFTSEDNKEEVSINVADEADVELELKVS
ELVACAPSKLSKNKINMDGMSRLVLEFLWYINLTKRIKIMLADAVNSTLEKIAS
IFDKLSNTAESPMGDETOEETLVDRKREKSTPDCYFVGVNFFDKENLEKASA
VCKIAGSKTLPAAGNFAIPVIGTISDVIISKIKATEVOVTKQCKSKCVLMTGQOS
OYPLMGROLVENYEVFRATLOSCLACKOEDYGVSLWELFNTENKILQITKHMOP
IMESFGITAOIMLSFGIYDXYLGHSVGEIACVINGIMTLEGLILYERKAMEN
IAGGALLAVORETADYLRKFEKVSVAIINSPKOVYFRAGTPELSDALAFKGGKOG
TFVNRORYPFHSLIDETHLVALRKCLADIRRGATRVLSVNTGQIINTFSEAYITK
TISAVFVCEVETLRNDGVAWVIDAGPAIILATFVKRIIVAEILSKHRIVOTCEKES
DVELTQSCLEKQSEFELIMPKYGGCTDDADLAHPVVOSEDTIONHEPELIEGR
LNGKIIIVAGAYOMKMDQLVKSKAGSDLVLTKNKFKVKKYIECKDAFLRWNSDMEI
ELVNSVTVCSAKVEMKEGALEIOVIGENKPEPVHFEITLYNGVSGSPRMD
SKRSDRCFESSLKPSFPAMPLIDSAMHSITRVSVPKRDSFTFLPVANGQVTLKQEDP
KIKINNOTVITSETDKFTOVNVYVACDQITGVKRMVTVLKLVAETPPTPTTA
KEAKSELEIYGFEDTTPYNOISTINNEWSAKLTNRVAKOPLOTNRNATTEBVALLD
LRTMDPEYEGIRPSEAGYLDQORLLTINWAKLIDLSISLPSKTGVFICGTYERS
QIVYAANOENPRAEMWGGTNSALAGRIAHMLKLGKGVLTVDTCSSSPALASACTT
IRSGCECAIVGTANVILHEMTSNVLONAOQVODPCFAPDVANGYKRSADACTLI
SKTGKIKSMASITVDMAGVNGNSSSLFTPNGISOLEIMKATKATSNRDIIEVPHCTC
KLDPIELSATSISLISIDCSISSSKNIGHAEKSGVLSCTILMSLQTYKKAQHL
KCPHEVITKGAALHQPVGEDLADHSHAFILNNGFPGSKCVLTAKPQNDPEIHESK
DVTPEMLSAHTIKSLERYISVFEFEYVASSASLRDMLSLDKVNHVQPLIYSK
KELVSDQDEDEKMP IQRQWVYKKEGTVEKATROVQLOPLPIYVNSILMKLNF
HDEIQGPCVPLREIPEFEKYLVEKAEKNNEPPIRVACVGRITLIPKITEMEVSCTIG
IIVFPLPNSLIEELKLTLEMSLVSPPOONSIVYACFNGKSHGEMTVGRSLASEKP
ILYEFVSDIHDVHLKKEFNODVFEISIFYSNGSKYVERLRITPTPOKMEHIFKCLIS
GGTGICKAVINLEIKPKKSVITTRKEPEKEETKYSLSVDADPTISKEDVHVFAGT
VKANSHINIKKIDILEENVLTKFOGAKNLNCCSLKGFYFSSSIACILGSEFGSNVTE
ANGVLSNFTLEWLSNFTKIIMGPKVEGMALKECCDIIKQIIMNOMIPLSNSDALSV
FSTLINSKHQVIVFDNFEIYVSRQHLNLSLEYILKTAQKQKQSPDESIFTYV
GIRMDNDRKHPFMDLIGIDSCMENLHTLNLHNLNLEANSIEEENATEKLNAYVAS
MLKSKSTESKSPKSDNITNRENDNVAVYIGASAEFGSEKNEEPEMENTLGGVATG
KENSILKSPGFGONKFNINDPAKMLDPOVKRFTIOHSYLALENSGVYKRDOLKRAY
FAGESEPDYCRPSQNDAMKRLVYMMNNNTIASASTCDLKGTSVSVSACSTTLVA
VANVKTIGSGSDVYALVGTASIAESECISGNSDKRTIESGCIKQVDRNSDEIV
EKGSGVGLVLCYSEAKKNDNTHFVLEKPAISNDGISRSEAPMAPDACQKQKMTVL
KGLQEDENRILEVECHATGTTLGDTIENNALDCAFCFKNLQIGCKNIGKAMVAVS
GLASLICKAKMLQNGKVPQVONSEFSDSGTFEPSTOKSFTONSFLTASISGFG
TNVMIIERIGIRBEKNDVDENVITLPSANSASLYEAKTMEVLOKSNLKLSTT
LHHRNOMNRAPIIASKITIGKVOITSFKITKSMSPKALFPAQGIQFANITLDE
LYAHNDEKSSLERCKTAHLGVSNLLEIIPYENYENEDLIQTLVLSQKALPVOCAL
FECIAKAVIKPOLIGHVSGVAAVYISOCISEEMMELIERTLSIQCKAMLAWV
NYEELPNSVIAIDIPITKCIYPTTEADLEDELEKFKICQVYKRNITKNHGSXKL
OPVOSKFNLCILSPSEPRPIPIISITGRIKRPDAEYCRKHLTAVNMLDVYNDIT
KSGVDVYETIGPVTGVLNLGKKNKSTIVIPCTGCKKPKASISDQIGELKCCGDIR
PLPPLKIDGNAGYCFDPYVORGSHKTPPVLYKFNYSKEMKSLNLSSTASVHIIQ
IYKNIGSTQTSKILYVHIDGSDSHONFVEKFLKQCYECLTITS25ASVHIIQ
LGLRCHQVTSRRIKIFVENFENLELKVVASN IADGCLYVRVTSGVGEHFLQISQ
NTHSIKTKREMNALVFGASGVGOSASRLHSGFVWIEVSSSKISSDIIRHLKVOE

```


VLKSLNLKHPDYVLANCVKSSSKJNKTFPEDEFTLTTRKYGNSLNKALDYNLKVD
KITTSSLSVPLDGNNDYASANCYEAELKSTPOKLVLSLPLVDSSRMKRETES
TKDMLSNLSNSELNDNLKIALTEETEGTLEFVSLDDPAQIARSLSTHREDEIEVP
ENKRECLVTSIRNIMKELTGSLIDISNPFSLGDSLSLQVWVOKLQKCTE
VNDLFNNPTLOOFTDFVOMNSQFSPCHODEKIPDFNNIQLFHSOTOMELRVLDV
TSRYNIIFELVYKDISFCWMSLKYSITSLIAYOTSYRSVENSRRPLOTOKVSLRESF
HNLDWDMLDLSOELVSHFELGOTPLRVRLADSDNNOHIVFNHOLLITDGM
TVLADITSLKLYTSFKLTVPLKSNQSAKLAIQOCVENLQAVNMAKSYFENEMHTI
IPIYDAVAEENHRPIKRIPIRMKIVDMKRYSTTHNCLAVFCVCSFPGOT
DLMFAANGNSNENSLIYFENFOMNALPFTSISEDEKEKILITVSKLQKRCNS
VPHFLVEONRKLNALISYFENFOMNALPFTSISEDEKEKILITVSKLQKRCNS
PGPILSIDYDSKVFESTVNFESTELKAKESFKAONNKLIHVRADSPFLO
IDLTSTWKPISLACSLIGASYPADVPSPEDQLYKSVNIYFESHVNEVNLK
GVTSKTPISLACSLIGASYPADVPSPEDQLYKSVNIYFESHVNEVNLK
SFKTRQVAGIYITFTSGSTGPKVFMAGQSVSLTSATKQCMREKRVRLVDK
VPEVSVNIIFGPMNGCCLISADHSTPTIDKLOCOYAPLPAVANGFISFEMRIAT
VETLTIGGETVSDALCALAKMFKMRIIOIYGPETCYWSTLNRCRISSENVAGNL
GESMENEKRYVNDISIRGLKVGISLARGVSNPKCGKNTSTGVDSKSGIYK
IGETDNOVKKGARDVTELEKELLCIGLOVAYISNOMLIAFLAGSTAFQYSDL
IAKLOKTHLPHRYVOTIDOMPLNNSCKIKSTLLOAFESVRSKSYKREMPKMSLJEK
IVASGDVGRKTEITDKFSDVGNLSLTAIQLAHLTELEKEDVKAHELLOSNGEDL
SKTLASVYKPKESKVPNVITKLEVPNSKFSIYLVAIGTTPYISFLQIPEKNTSL
YGIETDLKPCGDLRELAFYAOELIAHAGOKIIFVCHSGMIGMSREIVALEKNTSL
OIFPVMLEDSWLTRELDELIENTKOPITVFESLPDSEHRIDRAIKLAQILREKTSV
SDTKLYLFSKLOGLDAPKAVRVDLNEELSRMTCNGFDELIOPIETYLIDGPHES
CLKVENLKRYKDPILAFKRYPDQA"

18313..19760
/gene="G01D9.4"
join(18313..18387,18490..18569,18615..18737,18791..18886,
19060..19237,19285..19395,19444..19624,19684..19760)
/gene="G01D9.4"
/codon_start=1
/evidence=not_experimental
/protein_id="AA98700.1"
/db_xref="GI:1293794"
/translation="MTIFKTEETDTPKNAEGVKEENPKVVCGEADEEKHISNGWVP
MEGNFDRPEKRSKSSGSLILRLSPSPPLATPSPSPPLPPEVAVNARRRRKTKLTF
TLIAFSPTLVISFPLVQOGLISVAYOKGQIFVCHSGMIGMSREIVALEKNTSL
STKIDIPKFGSNRPAFLFDKQONLTAIVDYGNNGCFYKIDIRTKTIRSRMILNMLR
NIDIMAEVAVYOTRMVRETYNVGDELTKADISTFNTSLSRHCAFRQYTKLKQEAQ
TRSRAGMGPAISFASMGDSVQLEISF"
complement(19138..19236)
/note="Leu: codon recognized: CTA"
20621..21971
/gene="G01D9.3"
join(20621..20772,21464..21614,21664..21783,21834..21971)
/gene="G01D9.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AA98701.1"
/db_xref="GI:1293792"
/translation="MKTFNLSFLVISTYTPPCCRHLLGAACSKILMONTLFAKRC
NSDAEFLIOCCSCNTDVGMAVDILARSLVSHCHDRKGRPCDRLRSPPLSNTP
HNMTSCDGENPQIAFRTCRKSCGVCNVSVOYTLDNALQACRVQPAVEERKRVKRY
SSTTTTSSPSMEVINSTVOFMNHN"
complement(22269..22799)
/gene="G01D9.6"
complement(join(22269..22371,22415..22541,22588..22680,
22730..22799))
/gene="G01D9.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AA98702.1"
/db_xref="GI:1293793"
/translation="MIRPLTKVCLTSLIGVGVTVASTSMYSRRATPSEKWSITR
PRMFQNPBFKIGTIOKSTOLNPRALPCCDESGSICNLRRTDILKLTFOKQTEP
DPSLVCCNSCEAGISYRKRAQKQVNL"
32520..36916
/gene="G01D9.2"
join(32520..32678,33205..33505,33554..33662,33836..33936,
33983..34088,34698..34775,34824..35288,35332..35502,
35548..35847,36442..36637,36689..36916)
/gene="G01D9.2"

gene

Query Match 64.0%; Score 16; DB 3; Length 40031;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ttgctgttgaagcga 21
Db 7934 TCGCTGTTGAAGCGCA 7919

RESULT 29
AP002897/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AP002897 156425 bp DNA PLN 08-MAR-2001
Oryza sativa genomic DNA, chromosome 1, PAC clone: p0686E09.
AP002897
AP002897.2 GI:13365563
Oryza sativa (cultivar: Nipponbare) DNA, clone: p0686E09.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriatloideae; Oryzoideae; Oryza.
1 (sites)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone: p0686E09
Published only in Database (2000) In press
2 (bases 1 to 156425)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (19-OCT-2000) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@agr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Mar 16, 2001 this sequence version replaced gi:10944300.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI Nonredundant Protein database, or
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologues of the coding regions were searched against
NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID.

COMMENT

A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from Sp6 to T7 of the PAC clone.

This sequence of P0686E09 clone has an overlap with P0410E01 (DDBJ:AF002866) clone at the position 1 to 1233 of 5' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rpp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

Source

```
1. 156425
    /organism="Oryza sativa"
    /cultivar="Nipponbare"
    /db_xref="taxon:4530"
    /chromosome="1"
    /clone="P0686E09"
    join(4841..5341,5904..6008,6478..6591,7059..7241,
    7450..7578,7719..7781,8016..8111,9084..9173,9591..9707,
    9900..10022,10215..10307,10717..10878,10992..11114,
    11186..12178)
    /gene="P0686E09.1"
    join(4841..5341,5904..6008,6478..6591,7059..7241,
    7450..7578,7719..7781,8016..8111,9084..9173,9591..9707,
    9900..10022,10215..10307,10717..10878,10992..11114,
    11186..12178)
    /gene="P0686E09.1"
    /note="hypothetical protein
    similar to Arabidopsis thaliana chromosome 1, F26A9.2"
    /codon_start=1
    /protein_id="BAB39109.1"
    /db_xref="GI:13365564"
    /translation="MALNMTKLTQALAKASAVIEKTTTQVEVGTGRPLQDFELLEQ
    AGSGPGLAMRTYARPREGAPSAFVPSVWVLDKRLAEARAGITKAAADALD
    LARADARLVLKHPGLVHYVALDETKAMAMATEPVFASVANAALGCDLVNCKVE
    LKGVRESEMGLEVKHGLLOVSETDLELHNNALHRAISPEQTVFTLTKSGKLCGF
    GFALSVQATGSLSTQGFHSDYDVEDTALPLOSNTAPRLVSGSGKVSACDI
    FSPCLALHYARPLDCCNNKMYNNTITITSEFNTIPADLVVDIOMLSVA
    SRSAPAFSGSSFFRDTRKALRLFDHLLDERDMQKTEFLKALSDMKCFDSRV
    LKRVQVPLPCLAEFLRMVMQPMPLPVLTAISQDKGPELSTPLAIPVPTSAAG
    ETLILLVHADLILHKEHLISHLVPLMVARVDDTDRLOEVLKRTVLSROLDVVO
    LKQSVLPVHGLALKTTVAACVLIHVLVRYNALRLGLDVLPSLDEKLEILITL
    RCTAVDHTPTMCTGLGVANALIFKQCGVEFAAEVYIPLVPLTLTAHOGSPAKY
    ILFVKDITRKEKRGVTVNDCNTEKVSPLANGHSPLSRKCGQTEQIOTAKGA
    AMEDMKPTKASIPSSSDSNAMKQSPDFPSTOTKQPSALPDEFSTQCKPSSA
    IPQVTPATISPAISLPSLOSLAPSSGPTSGSCVPDIEMPPKSNISDFNALSVE
    QSKSGSLSDGLADIDPFDMPKASGASILLAVGHLSTSPQSIISGLSTGICFNS
    GSILOMKTOMMSAKPNTNLMVAVNSTGNVLMQNSALDFGCMPIGKRGSLSSSTV
    QGPPRLAPPSSAVSGRGRGOSALSRASRTPHGNSVSECPPIILDL"
    complement(join(13392..13540,13940..14123))
    /gene="P0686E09.2"
    complement(join(13392..13540,13940..14123))
    /gene="P0686E09.2"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB39110.1"
    /db_xref="GI:13365565"
    /translation="MHPSTRFAPYLAVPFRSGYRCRLIAAPAPCPRLQEPPIH
    NQCVAFALDPFGVARRPCLWLDGATVTEFEQPASSQLAKSLVCHSDVAEMPHM
    VHWTDAP"
    join(28972..29115,29117..29179)
    /gene="P0686E09.3"
    join(28972..29115,29117..29179)
    /gene="P0686E09.3"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB39111.1"
    /db_xref="GI:13365566"
    /translation="MHAAAASARTACSAARSTSLERWVGAIHGAGVVTLDQGMOR
    RAIPYRIOSAMMARAGRGREG"
    complement(join(29740..29837,30219..30380,32426..32546))
    /gene="P0686E09.4"
    complement(join(29740..29837,30219..30380,32426..32546))
    /gene="P0686E09.4"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB39112.1"
    /db_xref="GI:13365567"
```

```

    /translation="MRGATSLSPREVCVNLLERVLDGHSRTPQAIHLQNVLR
    KRCHDSTENALAAHKCGPELAVFTGLQAMTQASAPATSSREAOALKHRIPTRECV
    GQTEORTAKLKSRRCDART"
    join(33917..34076,35467..35816)
    /gene="P0686E09.5"
    join(33917..34076,35467..35816)
    /gene="P0686E09.5"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB39113.1"
    /db_xref="GI:13365568"
    /translation="MRWTLASAGVLDGRATHVLLTLRSKEROFTIIPALPFCFL
    GMYGLVAVLVKQNIETQAVRKCVCAVGVKMKAKSTYFVVGARSFMRPLEIT
    CVLSLIPSLSVFGKQILMFLGACVCSVRKAKLRDRAPCFPLSLAMFDGKIC
    NMHPFKIS"
    join(37532..38415,38988..39218,39280..39471,39552..39890,
    40001..40207,40381..40515,40601..40765)
    /gene="P0686E09.6"
    join(37532..38415,39005..39206,38988..39218,39280..39471,
    39552..39890,40001..40207,40381..40515,40601..40765)
    /note="hypothetical protein
    similar to Oryza sativa chromosome 1, P0665D10.16"
    /codon_start=1
    /protein_id="BAB39114.1"
    /db_xref="GI:13365569"
    /translation="MSEADRIVRNAALALDGGGAGVADKPAKADVFADLGSVPSP
    LKARASVATSSSSSSGSAKSPASNAAGLALAGSHSGELTASTPRLPGRRCGS
    GPLIFSGSGSGGGGDRGSTASTPTNLPGNICTPSGRVAVAAAPPPSRPD
    VLGSGTGVNGHSIMRGGGMAPASSITDSSSFLGHARSPATPAASASSGLOVY
    PTLGEMWKKKGHAALHYDRANVALCPESAACGNRRAALAGICRLADALRDCEAV
    RIDPANGRHSILAGLCRLDGTISAKRHLLTQAGLHLSDEKRLQDEVHOGSRIS
    DARKEVDMKSALREADAAIAGABSRTILTLRLGIMTSKARRHITQAGHLHOGPSB
    MEKIDVEMHOGSRIDARKVGDWKSALREDAALAAAGDSRLVMSLLATRSALLR
    LHKEIDESTLALIKLDSVLLRYRGAPSCMLAESYYSIVRAQVDMLGRCDEPDA
    VEAADNEFTIPGNAEVMILNNVKKVAKAOCNEIYKAKFSDASIAVSEGLKYPB
    SNPVYKCARACWGLKEWKEKAVDNCNARLTQPNYKALLRRASVAKWRADCYR
    DYEVLAKRLPDTVEAELPLTAQVALKTRREIDVSNMKGCEVEMTVQVDRADCS
    PGVIOFALSSGVSVYFSTIMNOOCTLTTPSVSLSCSECSLNFVLVQVVEDSPW
    KAEENRTVPFETIKDGVKVKEMICPSLALHRSVRYAVSSS"
    complement(join(47016..47264,47291..47363,48616..48748,
    48784..48832,49178..49324,49569..49684,50921..50967,
    51070..51107,51434..51652,51763..51861,51888..51986,
    52059..52112,52817..52933,53291..53532,54642..54855,
    55164..55345))
    /gene="P0686E09.7"
    complement(join(47016..47264,47291..47363,48616..48748,
    48784..48832,49178..49324,49569..49684,50921..50967,
    51070..51107,51434..51652,51763..51861,51888..51986,
    52059..52112,52817..52933,53291..53532,54642..54855,
    55164..55345))
    /gene="P0686E09.7"
    /note="hypothetical protein
    similar to Arabidopsis thaliana chromosome 3, MBK21.5"
    /codon_start=1
    /protein_id="BAB39115.1"
    /db_xref="GI:13365570"
    /translation="MSDREVSQTEENTWDMGVAREKKAADGLIAKVEGARSKCA
    LGQARHVIITLGRVQSGASRQPSAKTLASASAGSMDPFPYHGHGAAGGEGAA
    AAGYSYEVLDILARVYGRPLANPSSAALDLNLAIRMSGLVLIHQPTMGSHVYE
    QIEALYSNNTMTKRLPRLSSSLPTLPQRPGRKDAFYWMTPTCKFGSGCKDHPQWE
    GIPINMKSHLLEFALDNTLILYKGIINSCOHNVFLITLGIQOPREAVVESYPSF
    EGEPDCEPEMTCKCKFGSKCKRHKRPEKXNALAEVNTNMTNNSQKAPRSQVCSF
    MKLOKHLIADSSILIPYAKGCKCFKRMCKFPHKQDIETSSQNEPEASVTEGEDI
    GSADSVSAKQTTVAQAQEFNSKGLPMRCEVDYDCTPYMKSGCKRFGSTRTFNPDRL
    VLNPPLDGGTILPTPESMILNSSANPMQGFDAIAMPYMTGCFKDFDKCFHNP
    IDRSAPDSANMPEAESVOITLAGLRREDVAVCAFYMTGVCCKFGCGVQCFEPL
    RRENTILITOGRIKSMPSFLAYELNMKPHYLGSLYLRRLPULSESPSKAAAIISA
    QPVHAYGHLRLCAPOVAPLHPTAGCRPITRGSILATWSPRRAAGGDAVRLPGCLPV
    GCAMPARSAASGDHPLCCAGY"
    join(57612..57759,58207..58289,58617..58775,59266..59373,
    59488..59548,59653..59918,60187..60219)
    /gene="P0686E09.8"
    join(57612..57759,58207..58289,58617..58775,59266..59373,
```

59488. 59548.59653. .59918.60187. .60219)
 /gene="P0686E09.8"
 /note="contains ESTs D40230(S2054), D23027(C2064)
 hypothetical protein
 similar to Arabidopsis thaliana chromosome 1, F19K6.5"
 /codon_start=1
 /protein_id="BAB39116.1"
 /db_xref="GI:13365571"
 /translation="WVGWRAAGCARAVLRSLAAEAAMKQDGRFASYSSSGCVN
 APEGLQYANLRLAQAFASRGVALNFHOLIRNAGISTIRNLLAADAVPSSPLTP
 LDGEGTDKGAIVKRLKQAIKDKIKOSPKKVNIVAKLVKRMVEDALLOLVYVKR
 AKATYASRANAANHGIDPKLIVEEAFVKGGLYLKLSYHAGRCGVMPRCRL
 TVVREATNEEEAKTAKLKVSNYKKLTREKQLMHRLIEVSPMARRRKEEAGAAAL

Query Match 64.0%: Score 16; DB 8; Length 156425;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 atctgcgtttgaagc 18
 |||||
 Db 52932 ATCTGCTTTGAAGC 52917

RESULT 30
 AC016697/c 163489 bp DNA PRI 25-MAY-2001
 LOCUS Homo sapiens clone RP11-140C4, complete sequence.
 DEFINITION AC016697
 ACCESSION AC016697.8 GI:14196409
 VERSION HTG.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 163489)
 AUTHOR Waterston, R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 163489)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT Center project name: H_NH0140C04.
 FEATURES
 source
 1. 163489
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-140C4"
 BASE COUNT 46776 a 33045 c 34796 g 48872 t
 ORIGIN

Query Match 64.0%: Score 16; DB 9; Length 163489;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 gctgtttgaagcagc 22
 |||||
 Db 40201 GCTGTTTGAAGCGCAG 40186

RESULT 31
 AC023914/c 168177 bp DNA HTG 20-SEP-2000
 LOCUS

DEFINITION Homo sapiens clone RP11-44K20, WORKING DRAFT SEQUENCE, 22 unordered
 pieces.
 AC023914
 AC023914.3 GI:10198362
 HTG: HTGS_P14SE1: HTGS_DRAFT.
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 168177)
 AUTHOR Birren, B., Linton, L., Nusbaum, C., and Lander, E.
 TITLE Unpublished
 JOURNAL
 REFERENCE 2 (bases 1 to 168177)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., and Lander, E.
 TITLE Unpublished
 JOURNAL
 REFERENCE 2 (bases 1 to 168177)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bede, F., Boguski, L.,
 Boukhalter, B., Brown, A., Butte, G., Campiano, A., Castle, A.,
 Choquel, Y., Colangelo, M., Collins, S., Collumore, A., Cooke, P.,
 Dearlano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
 Fenestor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Landers, T., Laroque, K., Lehotzky, J., Levine, R.,
 Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M.,
 McEwan, P., McGuck, A., McKernan, K., McPherson, R., Meldrum, J.,
 Meneus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J., Naylor, J.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M.,
 Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tjelle, A.,
 Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
 Wu, X., Wyman, D., Ye, W.J., Young, G., Zahoun, J., Zimmer, A. and
 Zody, M.
 Direct Submission
 Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 20, 2000 this sequence version replaced gi:9102875.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: http://www-seq.wi.mit.edu
 Center code: MIBR
 Contact: sequence_submissions@genome.wi.mit.edu
 Project information
 Center project name: L1087
 Center clone name: 44_K_20
 ----- Summary Statistics
 Sequencing vector: M13: M77815; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 157528 bases at least Q40
 Consensus quality: 162841 bases at least Q30
 Consensus quality: 164870 bases at least Q20
 Insert size: 176000; agarose-fp
 Insert size: 166077; sum-of-ctnigs
 Quality coverage: 4.3 in Q20 bases; agarose-fp
 Quality coverage: 4.5 in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 22 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1263: contig of 1263 bp in length
 * 1264 1363: gap of 100 bp
 * 1364 2399: contig of 1036 bp in length

```

* 2400 2499: gap of 100 bp
* 2500 3713: contig of 1214 bp in length
* 3714 3813: gap of 100 bp
* 3814 27125: contig of 23312 bp in length
* 27126 27225: gap of 100 bp
* 27226 28854: contig of 1629 bp in length
* 28855 28954: gap of 100 bp
* 28955 31625: contig of 2671 bp in length
* 31626 31725: gap of 100 bp
* 31726 34084: contig of 2359 bp in length
* 34085 34184: gap of 100 bp
* 34185 36582: contig of 2398 bp in length
* 36583 36682: gap of 100 bp
* 36683 39629: contig of 2947 bp in length
* 39630 39729: gap of 100 bp
* 39730 43900: contig of 4171 bp in length
* 43901 44000: gap of 100 bp
* 44001 46837: contig of 2837 bp in length
* 46838 46937: gap of 100 bp
* 46938 52599: contig of 5662 bp in length
* 52600 52699: gap of 100 bp
* 52700 60235: contig of 7536 bp in length
* 60236 60335: gap of 100 bp
* 60336 67600: contig of 7265 bp in length
* 67601 67700: gap of 100 bp
* 67701 76219: contig of 8519 bp in length
* 76220 76319: gap of 100 bp
* 76320 85003: contig of 8684 bp in length
* 85004 85103: gap of 100 bp
* 85104 95875: contig of 10772 bp in length
* 95876 95975: gap of 100 bp
* 95976 108080: contig of 12105 bp in length
* 108081 108180: gap of 100 bp
* 108181 126996: contig of 18816 bp in length
* 126997 127096: gap of 100 bp
* 127097 145024: contig of 17928 bp in length
* 145025 145124: gap of 100 bp
* 145125 166554: contig of 21430 bp in length
* 166555 166654: gap of 100 bp
* 166655 168177: contig of 1523 bp in length.

```

FEATURES

```

source
  1..168177
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP11-44K20"
    /clone_id="RPC1-11 Human Male BAC"
    1..1263
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:left"
      1364..2399
        /note="assembly_fragment"
        2500..3713
          /note="assembly_fragment"
          3814..27125
            /note="assembly_fragment"
            27226..28854
              /note="assembly_fragment"
              28955..31625
                /note="assembly_fragment"
                31726..34084
                  /note="assembly_fragment"
                  34185..36582
                    /note="assembly_fragment"
                    36683..39629
                      /note="assembly_fragment"
                      39730..43900
                        /note="assembly_fragment"
                        44001..46837
                          /note="assembly_fragment"
                          46938..52599
                            /note="assembly_fragment"
                            52700..60235

```

```

/note="assembly_fragment"
60336..67600
/note="assembly_fragment"
67701..76219
/note="assembly_fragment"
76320..85003
/note="assembly_fragment"
85104..95875
/note="assembly_fragment"
95976..108080
/note="assembly_fragment"
108181..126996
/note="assembly_fragment"
127097..145024
/note="assembly_fragment"
145125..166554
/note="assembly_fragment"
166655..168177
/note="assembly_fragment"
clone_end:77
vector_side:right"
BASE COUNT 41092 a 41268 c 42342 g 41368 t 2107 others
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 2; Length 168177;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 3 atctgctgttgaagc 18
Db 81387 ATCTGCTGTTGAAGC 81372

```

```

RESULT 32
LOCUS AP003845
DEFINITION Oryza sativa chromosome 7 clone OJ1699_E05, *** SEQUENCING IN
ACCESSION AP003845.1 GI:14595190
VERSION AP003845
KEYWORDS HTG; HTGS-PHASE2.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1699_E05.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
Sasaki,T., Matsumoto,T. and Yamamoto,K.
1 (bases 1 to 171069)
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
clone:OJ1699_E05
Published Only in Database (2001) In press
2 (bases 1 to 171069)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (03-JUL-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@abrr.affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by
Combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers

```

FEATURES

```

FEATURES
  1..168177
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP11-44K20"
    /clone_id="RPC1-11 Human Male BAC"
    1..1263
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:left"
      1364..2399
        /note="assembly_fragment"
        2500..3713
          /note="assembly_fragment"
          3814..27125
            /note="assembly_fragment"
            27226..28854
              /note="assembly_fragment"
              28955..31625
                /note="assembly_fragment"
                31726..34084
                  /note="assembly_fragment"
                  34185..36582
                    /note="assembly_fragment"
                    36683..39629
                      /note="assembly_fragment"
                      39730..43900
                        /note="assembly_fragment"
                        44001..46837
                          /note="assembly_fragment"
                          46938..52599
                            /note="assembly_fragment"
                            52700..60235

```

```

source
1. .171069
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="7"
/clone="OJ169_E05"

BASE COUNT 47159 a 38616 c 38364 g 46546 t 384 others
ORIGIN

Query Match 64.0% Score 16; DB 2; Length 171069;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctcgtgttgaagc 18
|||||
Db 123035 ATCTGCTGTGAGAC 123050

RESULT 33
LOCUS AC090534 191677 bp DNA HTG 28-JUL-2001
DEFINITION Mus musculus chromosome 18 clone RP23-233F17 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 23 unordered pieces.
ACCESSION AC090534 7 GI:15027685
VERSION AC090534
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 191677)
Li, L., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Halder, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Pereira, A., Shim, C., Thomas, E. and Kuchelapatil, R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 191677)
Grills, G., Li, L., Montgomery, K.T., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Halder, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Pereira, A., Shim, C., Thomas, E. and Kuchelapatil, R.
Direct Submission
Submitted (02-MAR-2001) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Jul 28, 2001 this sequence version replaced gi:14993657.

-----Genome Center
Center: Albert Einstein College of Medicine
Center Code: AECOM
Web site:
http://sequence.aecom.yu.edu/cgi-
bin/ws.exe/mouseDB/mouseSeq/mouseseqtable.htm
Contact: htgs@sequence.aecom.yu.edu
-----Summary Statistics
Center project name: AJT
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 223066 at least Q20
*Consensus quality: 213794 at least Q30
*Consensus quality: 199611 at least Q40
*Estimated insert size: agarose-FP - N/A
**Estimated insert size: 191237 - sum-of-coverage
Quality coverage: agarose-FP - N/A
Quality coverage: 6.1 x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

FEATURES
Source
1. .171067
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="18"
/clone="RP23-233F17"
/sex="male"
1. .42624
/clone="assembly_name:Contig108"
vector_side:left"
vector_side:right"
160280..160617
/clone="assembly_name:Contig104"
160638..170066
/clone="assembly_name:Contig103"
170087..177793
/clone="assembly_name:Contig102"
177814..179001
/clone="assembly_name:Contig101"

```

```

misc_feature 179022.180140
/Note="assembly_name:Contig100"
misc_feature 180161.180415
/Note="assembly_name:Contig99"
misc_feature 180436.180872
/Note="assembly_name:Contig98"
misc_feature 180893.181101
/Note="assembly_name:Contig97"
misc_feature 181122.182014
/Note="assembly_name:Contig96"
misc_feature 182035.183420
/Note="assembly_name:Contig95"
misc_feature 183441.185321
/Note="assembly_name:Contig94"
misc_feature 185342.185549
/Note="assembly_name:Contig93"
misc_feature 185570.186845
/Note="assembly_name:Contig92"
misc_feature 186866.187015
/Note="assembly_name:Contig91"
misc_feature 187036.187715
/Note="assembly_name:Contig90"
misc_feature 187736.188117
/Note="assembly_name:Contig89"
misc_feature 188138.189263
/Note="assembly_name:Contig88"
misc_feature 189284.190807
/Note="assembly_name:Contig87"
misc_feature 190828.191677
/Note="assembly_name:Contig86"
BASE COUNT 52515 a 43754 c 41414 g 53467 t 527 others
ORIGIN

```

```

Query Match 64.0% Score 16; DB 2; Length 191677;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggaatgctgtttgaa 16
|||||
Db 156118 GCATCTGCTTTTGA 156133

```

```

RESULT 34
AC007104/c
LOCUS
DEFINITION Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 16
unordered pieces.
ACCESSION AC007104
VERSION AC007104.4 GI:5523795
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 200000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 200000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1999) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304 USA
On Jul 17, 1999 this sequence version replaced gi:4757683.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
COMMENT

```

```

* 1
* 1370 1369: contig of 1369 bp in length
* 3775 3774: gap of unknown length
* 4879 4878: contig of 1104 bp in length
* 7284 7283: gap of unknown length
* 8556 8555: contig of 1272 bp in length
* 10961 10960: gap of unknown length
* 12208 12207: contig of 1247 bp in length
* 14613 14612: gap of unknown length
* 16263 16262: contig of 1650 bp in length
* 16263 16267: gap of unknown length
* 18668 20176: contig of 1509 bp in length
* 20177 22581: gap of unknown length
* 22582 24278: contig of 1697 bp in length
* 24279 26683: gap of unknown length
* 26684 28198: contig of 1515 bp in length
* 28199 30604: gap of unknown length
* 30604 32359: contig of 1756 bp in length
* 32360 34764: gap of unknown length
* 34765 39958: contig of 5194 bp in length
* 39959 42363: gap of unknown length
* 42364 57085: contig of 14722 bp in length
* 57086 59490: gap of unknown length
* 59491 74905: contig of 15415 bp in length
* 74906 77310: gap of unknown length
* 77311 84273: contig of 6963 bp in length
* 84274 86678: gap of unknown length
* 86679 118061: contig of 31383 bp in length
* 118062 120466: gap of unknown length
* 120467 152965: contig of 32499 bp in length
* 152966 155370: gap of unknown length
* 155371 200000: contig of 44630 bp in length.
FEATURES
source
1..200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"

```

```

BASE COUNT 38319 a 42757 c 43367 g 37466 t 38091 others
ORIGIN

```

```

Query Match 64.0% Score 16; DB 2; Length 200000;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 arcctgctgttgaagc 18
|||||
Db 129264 ATCTGCTGTGTAAGC 129249

```

```

RESULT 35
AX120085/c
LOCUS
DEFINITION Sequence 1 from Patent EP1108790.
ACCESSION AX120085 AX114121
VERSION AX120085.1 GI:14036800
KEYWORDS
SOURCE
ORGANISM Corynebacterium glutamicum.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 349980)
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 1 20-JUN-2001;
KYOMA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source
1..349980
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
/Note="Seq 1 to long (3,309,400) split in 11, seq 1
0.000.001 0.349.980 349.980"

```

BASE COUNT 79703 a 91547 c 98381 g 80349 t
ORIGIN

Query Match 64.0%; Score 16; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ctgttgaagcagc 23
|||||
Db 238128 CTGTTGAAGCGCAGC 238113

RESULT 36
LOCUS PPHE7D 306 bp DNA VRL 10-FEB-1999
DEFINITION Human papillomavirus type 21 DNA for E7 protein, complete cds.
ACCESSION D50548
KEYWORDS E7 protein.
SOURCE Human papillomavirus type 21 DNA.
ORGANISM Human papillomavirus type 21
Virus; dsDNA viruses, no RNA stage: Papillomaviridae;
Papillomavirus.
1 (bases 1 to 306)
REFERENCE
AUTHORS Kiyono,T.
TITLE Direct Submission
SUBMITTED (11-MAY-1995) to the DDBJ/EMBL/Genbank databases. Tohru
Kiyono, Aichi Cancer Center, Research Institute, Laboratory of
Viral Oncology; 1-1 Kanokoden, Chikusa-ku, Nagoya, Aichi 464, Japan
(E-mail: H4714@nuc.cc.nagoya-u.ac.jp, Tel:052-762-6111(ex.8838),
Fax:052-763-5233)
2 (bases 1 to 306)
REFERENCE
AUTHORS Adachi,A., Kiyono,T., Ohashi,M. and Ishibashi,M.
TITLE Detection of human papillomavirus type 47 DNA in malignant lesions
of epidermodysplasia verruciformis with protocols for precise
typing of the related HPV DNAs
UNPUBLISHED (1995)
JOURNAL 3 (sites)
REFERENCE
AUTHORS Adachi,A., Kiyono,T., Hayashi,Y., Ohashi,M. and Ishibashi,M.
TITLE Detection of human papillomavirus (HPV) type 47 DNA in malignant
lesions from epidermodysplasia verruciformis by protocols for
precise typing of related HPV DNAs
J. Clin. Microbiol. 34 (2), 369-375 (1996)
96381005

JOURNAL MEDLINE
FEATURES
source Location/Qualifiers
1..306
/organism="Human papillomavirus type 21"
/db_xref="taxon:31548"
1..306
/gene="E7"
1..306
/gene="E7"
/note="putative"
/codon_start=1
/product="E7 protein"
/protein_id="BAA09116.1"
/db_xref="GI:808884"
/translation="MIGKEVTLQDVLVLENEQPVQVVDLFCREELPSDOQTEFEEL
PERTSYKVVTPGCGCKVRILFVNATRAIRFQMLFEELQLLCPGCGCKHGGCS"
BASE COUNT 88 a 59 c 81 g 78 t
ORIGIN

Query Match 60.0%; Score 15; DB 14; Length 306;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctgctgttgaag 17
|||||
Db 236 ATCTGCTGTTGAAG 250

RESULT 37
LOCUS PEU244276 625 bp DNA PLN 11-MAY-1999
DEFINITION Pringsheimia euphorbiae 5.8S rRNA gene and internal transcribed
spacers 1 and 2 (ITS1, ITS2), strain CBS 747.71.
ACCESSION AJ244276
KEYWORDS AJ244276.1 GI:4808321
5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer 1;
internal transcribed spacer 2; ITS1; ITS2.
pringsheimia euphorbiae
pringsheimia euphorbiae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et
Chaetothyriomycetes incertae sedis; Dothioraceae; Pringsheimia.
1 (bases 1 to 625)
REFERENCE
AUTHORS Yurlova,N., Hoog,G.D. and Gerrits van den Ende,A.
TITLE Taxonomy of Aureobasidium and allied genera
UNPUBLISHED
2 (bases 1 to 625)
REFERENCE
AUTHORS Hoog,G.D.
TITLE Direct Submission
SUBMITTED (03-MAY-1999) Hoog G.D., Research, Centraalbureau voor
Schimmelcultures, P.O.Box 273, 3740 AG Baarn, NETHERLANDS
Location/Qualifiers
1..625
/organism="Pringsheimia euphorbiae"
/strain="CBS 747.71"
/db_xref="taxon:93493"
<1..196
/note="internal transcribed spacer 1, ITS1"
197..351
/gene="5.8S rRNA"
/product="5.8S ribosomal RNA"
197..351
/gene="5.8S rRNA"
352..>625
/note="internal transcribed spacer 2, ITS2"
BASE COUNT 170 a 128 c 135 g 192 t
ORIGIN

Query Match 60.0%; Score 15; DB 8; Length 625;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gtttgaagcagca 24
|||||
Db 438 GTTGAAGCCGACCA 452

RESULT 38
LOCUS AX110623 810 bp DNA PAT 30-APR-2001
DEFINITION Sequence 1356 from Patent WO0123604.
ACCESSION AX110623
VERSION AX110623.1 GI:13926915
KEYWORDS
ORGANISM
SOURCE
Serriella marcescens.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Serratia.
1 (bases 1 to 810)
REFERENCE
AUTHORS Bergeron,M.G., Boissinot,M., Huletsky,A., m Nard,C., Ouellette,M.,
Picard,F.J. and Roy,P.H.
TITLE Highly conserved genes and their use to generate probes and primers
for detection of microorganisms
JOURNAL Patent: WO 0123604-A 1356 05-APR-2001;
Infecio Diagnostica (I.D.I.) INC. (CA)
FEATURES
source Location/Qualifiers
1..810
/organism="Serriella marcescens"
/db_xref="taxon:615"
BASE COUNT 129 a 270 c 268 g 143 t
ORIGIN

Query Match 60.0%; Score 15; DB 6; Length 810;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ctgtttgaagcgag 22
 |||||

DB 742 CTGTTGAAGCGCAG 756

RESULT 39
 AF128956 1551 bp DNA BCT 21-JUN-1999
 LOCUS Escherichia coli strain E2348/69 flagellin (fliC) gene, partial
 DEFINITION cds
 ACCESSION AF128956 GI:5107854
 VERSION AF128956.2 GI:5107854
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 1551)
 AUTHORS Reid, S.D., Selander, R.K. and Whittam, T.S.
 TITLE Sequence diversity of flagellin (fliC) alleles in pathogenic
 Escherichia coli

JOURNAL J. Bacteriol. 181 (1), 153-160 (1999)
 MEDLINE 99084952

REFERENCE 2 (bases 1 to 1551)
 AUTHORS Reid, S.D. and Whittam, T.S.
 TITLE Direct Submission
 SUBMITTED (16-FEB-1999) Biology Department, Institute of Molecular
 Evolutionary Genetics, The Pennsylvania State University, 208
 Mueller Laboratory, University Park, PA 16802, USA

REFERENCE 3 (bases 1 to 1551)
 AUTHORS Reid, S.D. and Whittam, T.S.
 TITLE Direct Submission
 SUBMITTED (21-JUN-1999) Biology Department, Institute of Molecular
 Evolutionary Genetics, The Pennsylvania State University, 208
 Mueller Laboratory, University Park, PA 16802, USA

REMARK
 COMMENT Sequence update by submitter
 FEATURES
 location/Qualifiers
 1..1551
 /organism="Escherichia coli"
 /strain="E2348/69"
 /db_xref="taxon:562"
 /note="enteropathogenic"
 <1..>1551
 /gene="fliC"
 <1..>1551
 /gene="fliC"
 /codon_start=1
 /transl_table=11
 /product="flagellin"
 /protein_id="AAD28527.2"
 /db_xref="GI:5107854"
 /translation="OSALSSSLERLSSGRLINSKADDAAGAIANRFTSNIKGLTQAA
 RANAGISVAOTTEGALSEINNNLORITREITVOAQTGNSDSDIDSIDETIKSRDEI
 DVGSGTOBNGVNLAKGSKTIQYGANDGOTITIDLRKIDSDTLGLGANGFVNNKGET
 ANTAAATLRKMSGFTAAAPGAGTGVOTYTDKSAVSVNIIINAVAGANGFVNTSADV
 GGTTPAAATYVYKNDTNSYSASDDISANIAAFINPOARDPTKATVTIGKDDQVY
 IDKSGNLTAAADGAVLYMDATNLTNNAGGDTOTLAKVAIATGAKAATIGDKGTF
 TSDGTAFDGASMSIDANTFANAKNDTATGATGAKTYSVTGSAADATVAMSNVLSI
 TPTTYAAGDGSITTTEDAAAGKLYKSGDKLTPTTSKAESESTDPALADDIISOI
 DKERSIGAVGNRILDSATVNLNNTTNLFEAOSRIODADYATVAEVSNMSKAOITDOAGN
 SYLAKANOV"

BASE COUNT 443 a 375 c 368 g 365 t

Query Match 60.0%; Score 15; DB 1; Length 1551;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ctgtttgaagcgag 22
 |||||

DB 1432 CTGTTGAAGCGCAG 1446

RESULT 40
 SMAAC3VB 1572 bp DNA BCT 26-APR-1993
 LOCUS Serratia marcescens N-acetyltransferase (aac3-Vb) gene, complete
 DEFINITION cds.
 ACCESSION M97172.1 GI:152811
 VERSION M97172.1 GI:152811
 KEYWORDS acetyltransferase.
 SOURCE Serratia marcescens DNA.
 ORGANISM Serratia marcescens
 Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Serratia.

REFERENCE 1 (bases 1 to 1572)
 AUTHORS Rather, P.N., Mierzwa, R., Hare, R., Miller, G. and Shaw, K.
 TITLE Cloning and DNA sequence analysis of an aac(3)-Vb gene from
 Serratia marcescens

JOURNAL Antimicrob. Agents Chemother. 36, 2222-2227 (1992)
 MEDLINE 93073766

FEATURES
 location/Qualifiers
 1..1572
 /organism="Serratia marcescens"
 /db_xref="taxon:615"
 656..1465
 /gene="aac3-Vb"
 656..1465
 /gene="aac3-Vb"
 /codon_start=1
 /transl_table=11
 /product="acetyltransferase"
 /protein_id="AAA26548.1"
 /db_xref="GI:152812"
 /translation="MNTIESITADLHGLGVRPGDILMVHSLKAVGFEVGGASVYSA
 LRAAVGSAGTLMGYASWDRSPYEETLNCARMEDEFRLRRMPDLATGTVPGGLNLR
 FLLEAPDAPRSGAPDASWAVAGPLAATLTPHRLGOLAGSPFLERGVGCKVLLIG
 APLDSVTVIHAELIAPLIPKRRVYVMPMLGPDGRVRLMELADEPDSNGILDCFAVDG
 KPDVETITAAVYVELGHRRESIVRAFSYIFPADIVSFGVTYILEQHPGAP"

BASE COUNT 276 a 516 c 506 g 274 t

Query Match 60.0%; Score 15; DB 1; Length 1572;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ctgtttgaagcgag 22
 |||||

DB 1397 CTGTTGAAGCGCAG 1411

RESULT 41
 AF239157 1616 bp mRNA ROD 06-JUL-2001
 LOCUS Rattus norvegicus DEXRAS1 (Dexas1) mRNA, complete cds.
 DEFINITION AF239157
 ACCESSION AF239157
 VERSION AF239157.1 GI:7230767
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 1616)
 AUTHORS Pang, M., Jaffrey, S.R., Sawa, A., Ye, K., Luo, X. and Snyder, S.H.
 TITLE Dexas1: a G protein specifically coupled to neuronal nitric oxide
 synthase via CAPON

JOURNAL Neuron 28 (1), 183-193 (2000)
MEDLINE 20537828
PUBMED 11086993
REFERENCE 2 (bases 1 to 1616)
AUTHORS Fang, M., Jaffrey, S. R., Sawa, A., Ye, K. and Snyder, S. H.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-2000) Neuroscience, Johns Hopkins University, 725 N. Wolfe St, Baltimore, MD 21205, USA
FEATURES
source location/Qualifiers
1..1616
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
gene 1..1616
/gene="Dextrast"
118..960
/gene="Dextrast"
/note="Ras-related protein"
/codon_start=1
/product="DEXRAS1"
/protein_id="AAF3090.1"
/db_xref="GI:7230768"
LOCUS AF009246 1623 bp mRNA ROD 13-FEB-1998
DEFINITION Mus musculus ras-related protein (DEXRAS1) mRNA, complete cds.
ACCESSION AF009246
VERSION AF009246.1 GI:2253712
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1623)
AUTHORS Kempainen, R. J. and Behrend, E. N.
TITLE Dexamethasone rapidly induces a novel ras superfamily member-related gene in Atr-20 cells
JOURNAL J. Biol. Chem. 273 (6), 3129-3131 (1998)
MEDLINE 98123070
REFERENCE 2 (bases 1 to 1623)
AUTHORS Kempainen, R. J. and Behrend, E. N.
TITLE Direct Submission
JOURNAL Submitted (18-JUN-1997) Physiology & Pharmacology, Auburn University College of Veterinary Medicine, 213 Greene Hall, Auburn, AL 36849, USA
FEATURES
source location/Qualifiers
1..1623
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="corticotrope tumor cell line Atr-20"
gene 1..1623
/gene="DEXRAS1"
142..984
/gene="DEXRAS1"
/note="induced by dexamethasone"

/codon_start=1
/product="ras-related protein"
/protein_id="AAC53538.1"
/db_xref="GI:2253713"
/translation="MKLAAMIKKMPDSELSIPAKNCYRMVILGSSKVGKTAIVSRF
LNGREDAVPTIEDHRRFYSIRGEVQDLDTSGNHPFAPARRLSILTGDFILV
FSLNDRDSEFEVORLKOIILDTKSLCKNKTENVDPVLTICGNKGRDPEYREVEOREL
EOLVGDPPORCAVFEISAKKNSSLDPMRALFPAKLPSEMSPLDKRVSVQCDVLH
KKALRNKKLLRAGSGGGDHGDAFGILAPFAPRRSVHSDLMYIREKTSVSGAKDKER
CVIS"
BASE COUNT 412 a 437 c 417 g 357 t
ORIGIN
Query Match 60.0%; Score 15; DB 10; Length 1623;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ggaatcactgttca 15
Db 505 GGATCTGCTTTGA 491
RESULT 43
AF039572
LOCUS AF039572 2795 bp DNA BCT 02-DEC-1998
DEFINITION Serratia liquefaciens serrawettlin synthase (swrA) gene, partial
cgs
ACCESSION AF039572
VERSION AF039572.1 GI:2760927
KEYWORDS Serratia liquefaciens.
SOURCE Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
ORGANISM Serratia.
REFERENCE 1 (bases 1 to 2795)
AUTHORS Lindum, P. W., Anthoni, U., Christophersen, C., Eberl, L., Molin, S. and Givskov, M.
TITLE N-Acyl-L-homoserine lactone autoinducers control production of an extracellular lipopeptide biosurfactant required for swarming motility of Serratia liquefaciens MGI
JOURNAL J. Bacteriol. 180 (23), 6384-6388 (1998)
MEDLINE 99047587
REFERENCE 2 (bases 1 to 2795)
AUTHORS Lindum, P. W., Eberl, L., Anthoni, U., Christophersen, C., Molin, S. and Givskov, M.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1997) Dept. of Microbiology, Tech. University of Denmark, Bldg. 301, Lyngby DK-2800, Denmark
FEATURES
source location/Qualifiers
1..2795
/organism="Serratia liquefaciens"
/strain="MGI"
/db_xref="taxon:614"
111..>2795
/gene="swrA"
111..>2795
/gene="swrA"
/function="produces the lipodispeptide serrawettlin W2"
/note="serrawettlin W2 is required for the rapid surface translocation known as swarming motility."
/codon_start=1
/transl_table=1
/product="serrawettlin synthase"
/protein_id="AAC82333.1"
/db_xref="GI:2760928"
/translation="MLPWRKSNCKLPPIQRELMASNSGSEGLPFNNMGYGTIEG
PLQIALRQALDAIIEETPAMQAHPTODELEYOYARSHDPLSAHDSQSRPELA
AQRMWRQDADRVISALDQDLGFALTLTAGSTILTRRHIMLVDRSAEFPYRAL
VYNALAEKKSIPESNCSFSRLYEHDIYDRSSRFATDKQWOGYTLAAHSLTROK
LVPYSQMOISNHTLTINEIAIDRGAECAGVHAHILAAVAALCFHSITGQTLNLSL
PVLTGRNSTGMTANVPLTILAIINPSRIADFAQOVAEIAKVRVROLRGEDIDRD
RSTMSSMPGPAINIVSFDHGDPPWCKTMYGCGNLAVSDMGILFEYEDQQAQLDVM

BASE COUNT 545 a 786 c 888 g 576 t
 ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 2795;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ctgtttgaagcgag 22
 |||
 Db 2571 CTGTTGAAGCGCAG 2585

RESULT 44
 LOCUS PMO238607 6043 bp RNA VRL 15-OCT-1999
 DEFINITION Potato mop-top virus RNA1 complete genome, Swedish isolate (Sw).
 ACCESSION AJ238607
 VERSION AJ238607.1 GI:6066602
 KEYWORDS polymerase.
 SOURCE Potato mop-top virus.
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Furoviruses.
 REFERENCE 1 (bases 1 to 6043)
 Savenkov E.I., Sandgren, M. and Valikonen, J.P.
 Complete sequence of RNA 1 and the presence of tRNA-like structures
 in all RNAs of Potato mop-top virus, genus Pomovirus
 J. Gen. Virol. 80 (Pt 10), 2779-2784 (1999)
 JOURNAL MEDLINE 20037846
 REFERENCE 2 (bases 1 to 6043)
 Savenkov E.I.
 Direct Submission
 Submitted (27-APR-1999) Savenkov E.I., Plant Biology, Slu, box
 7080, se-750 07, Uppsala, SWEDEN
 JOURNAL TITLE location/Qualifiers
 FEATURES
 source 1. 6043
 /organism="Potato mop-top virus"
 /viralon
 /isolate="Swedish (Sw)"
 /db_xref="taxon:37128"
 /country="Sweden"
 /note="RNA1 RNA; genus Pomovirus"
 115. 4026
 /function="replication"
 /note="ORF1, readthrough protein"
 /product="polymerase"
 /protein_id="CAB58364.1"
 /db_xref="GI:6066603"
 /translation="MDQSLKTSDDYVNAIIHSAFPVSGDLINAKNVIVDDIKD
 STENKKKKKKIVKRNLDNDQILLAELEPERRVYSSVHRTSHMAARKIETDVI
 FTSFKNQVITIDGNMATBAKDDGGEVHCPCPLIDFDADARKMTLIDFNKPTDIA
 KVASAKAAVAADIKKCDLISENAKADYDANDLNGTFCQNKFEEDCYDHSSTGDC
 KKVAGMAHSTYDILVDLVASAMERKKVRLGTFLFSADILLICKREELSVNGFY
 IIDGSEIKYSFYDDPNCGEYHNLSIMLYVTQFVKAAGAVYIELTEMRGDMFT
 ITDASERAVGVLDHDSSTCKLPINKRDIVPPLPIDRATDDELVREELISREYNRA
 LKAPOLKNOVTAEGILISYASTANNAVITGSAKTSKVDPKLIPMTITTLVYQF
 LOKAKORVLGKLKSKVEELTSLGTLISVHRVEROSLYVORGUGVAKMWOYVQF
 DLVGHVPLYLEINDRIKLSALNVNGFSLSELEKYSLVLEEERQRISEI
 VSEKIQILGEGYVKGESLKPQKAVSGDGAQWVSGECLILYNTAKCEKVEKPR
 FSKVYLEWISGNCFALNNNSGDEVHMFDSLKACGKRAMKROVSAVTEFSDAEYDNI
 ENTNEELIEEVEVITKTEBOEDMDPIGNLPDVPDSDASAGVCLTESASTSDEDCD
 MDLIIVACERAFATIKKEEYVMDQVOSVEIETLYRDEENHAMLSGESSTSSOSQ
 ELVAVGSVPLSKMAQWEDSEVRARQCAVDHDLSDMEVKYAKMPEPEADDDERTK

BASE COUNT 1791 a 935 c 1578 g 1739 t
 ORIGIN

Query Match 60.0%; Score 15; DB 14; Length 6043;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ctgtttgaagcgag 22
 |||
 Db 4645 CTGTTGAAGCGCAG 4631

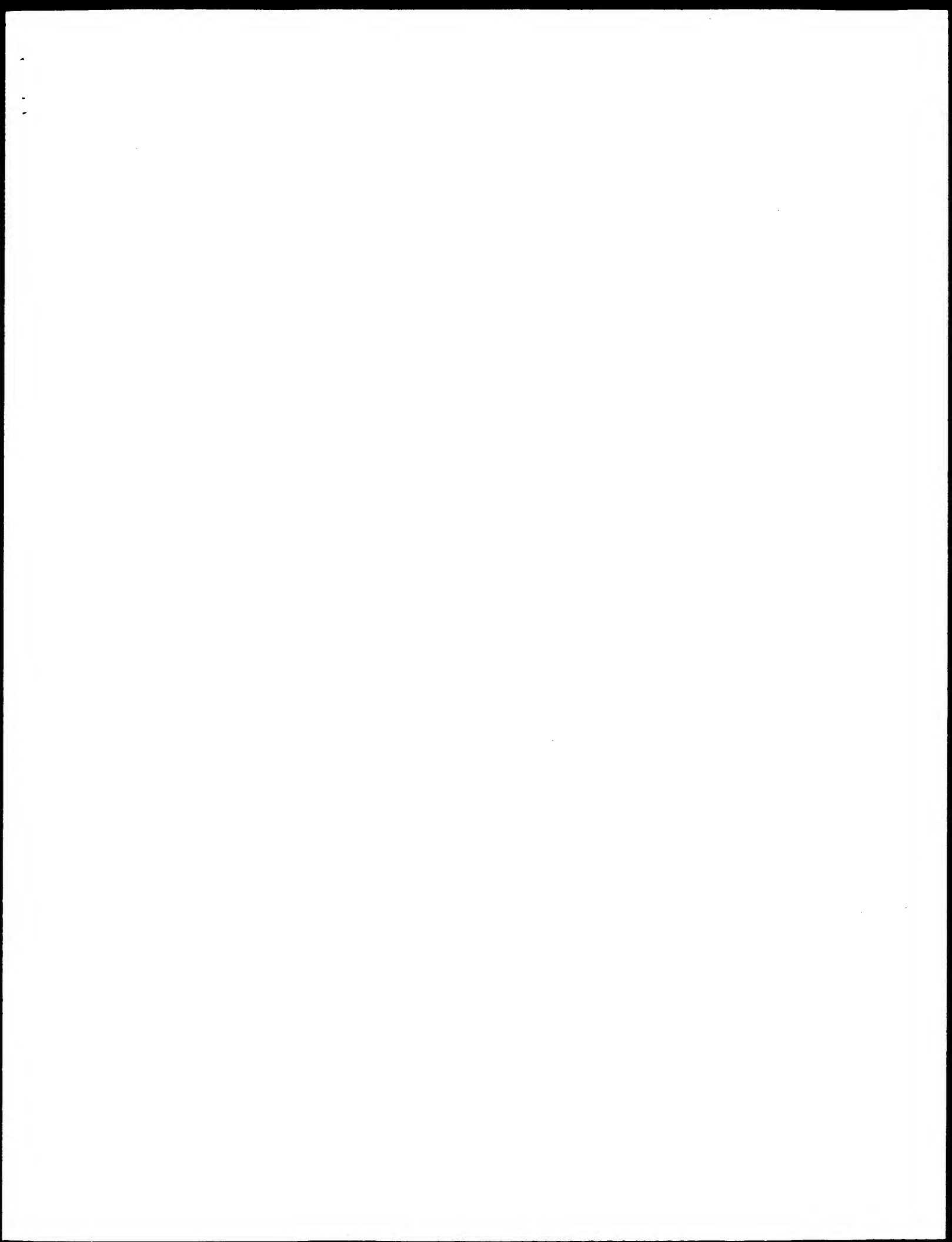
RESULT 45
 LOCUS HPJ31779 7779 bp DNA VRL 17-OCT-1995
 DEFINITION Human papillomavirus type 21, complete genome.
 ACCESSION U31779
 VERSION U31779.1 GI:1020170
 KEYWORDS
 SOURCE
 ORGANISM Human papillomavirus type 21.
 Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 Papillomavirus.
 REFERENCE 1 (bases 1 to 7779)
 Dellius H.
 Sequenced by Hajo Dellius, Deutsches Krebsforschungszentrum,
 Angewandte Tumorstudiologie, I.N.F. 506, W-6900 Heidelberg, Germany
 Unpublished
 JOURNAL MEDLINE 2 (sites)
 REFERENCE 2 (sites)
 Kremsdorff, D., Favre, M., Jablonska, S., Obalek, S., Rueda, L.A.,
 Lutzner, M.A., Blanchet-Bardon, C., Van Voort, Vader, P.C. and Orth, G.
 Molecular cloning and characterization of the genomes of nine newly
 recognized human papillomavirus types associated with
 epidermodyplasia verruciformis
 J. Virol. 52 (3), 1013-1018 (1984)
 JOURNAL MEDLINE 85033930
 REFERENCE 3 (sites)
 Kiyono, T., Hiraiwa, A. and Ishibashi, M.
 Differences in transforming activity and coded amino acid sequence
 among E6 genes of several papillomaviruses associated with
 epidermodyplasia verruciformis
 Virology 186 (2), 628-639 (1992)
 JOURNAL MEDLINE 92124737
 REFERENCE 4 (bases 1 to 7779)
 Farmer, A.D.
 Direct Submission
 Submitted (18-JUL-1995) Andrew D. Farmer, HIV Sequence Database,
 Los Alamos National Laboratory, W-10, Mail Stop K710, Los Alamos,
 NM 87501, USA
 HPV21 was originally isolated from skin warts of an
 epidermodyplasia verruciformis (EV) patient [2]. Hybridization
 assays and phylogenetic reconstructions based on DNA sequences
 indicate that HPV21 is most closely related to HPV14 and HPV20, and
 then to HPV19 and HPV25. This grouping agrees with assays of the
 degree of transforming activity of the E6 gene (these related HPV
 types had relatively low transforming activity as compared to HPVs
 5, 8, and 47), and clustering of similarity of amino acids in the
 second zinc finger domain of E6 [3]. The E6 gene of HPVs 14, 21,
 and 25 can enhance the induction of anchorage independent growth of
 3T1 cells by the HPV16 E7 gene, although again less effectively
 than that of HPVs 5, 8, and 47.
 location/Qualifiers

source
1. .7779
/organism="Human papillomavirus type 21"
/db_xref="taxon:31548"
/note="Cloned HPV-21 was obtained from the Papillomavirus Reference Center, Heidelberg and sequenced by Dr. H. Delius"
gene
200. .706
/gene="E5"
200. .706
/gene="E6"
/note="putative"
/codon_start=1
/product="transforming protein E6"
/protein_id="AA079394.1"
/db_xref="GI:1020171"
/translation="MADSDTSDADEGSPKRRRLLEENTSSLEPPPLPATIRLNL
EIPDLDCVPCNFCGNPLTHLECEFDKELSLMKDHCVFACRCVCAATATYNE
FYESTVGRDIEITKSIFFIDIVRCYCNMKFLDSLEKLDICGRKFFPKVRSMKGI
CRCKHFO"
706. .1011
/gene="E7"
706. .1011
/gene="E7"
/note="putative"
/codon_start=1
/product="transforming protein E7"
/protein_id="AA079395.1"
/db_xref="GI:1020172"
/translation="MIGKEVTLDDIVLELNELOPEVQVVDLFCSEELPSSEQTEEL
PERTARKVTPGCCCKVLRIFVNATQFAIRFQNLFEELQLCEPCGNCNKGGS"
998. .2809
/gene="E1"
998. .2809
/gene="E1"
/note="putative"
/codon_start=1
/product="replication protein E1"
/protein_id="AA079396.1"
/db_xref="GI:1020173"
/translation="MADPKGTSKEGIEEMCIYEAESQVENDLEELPDRDSDISE
LINDLEQNSRELPHQESKSESEBOLKRRKYLSPKAVADLSPLSITLSPQK
SKRLRVEDSGLECLTINEEDVSEVEPALDSQVAEQIGTVDIHKELLRASNN
KALIMAKFEFFGENDLROKRSYCCCNANVLAVYAVHDLLESSKOLQOHCY
IITGIGAMSLFLCEKVGKNGTIVHKLMTAMLVHEKOTISEPKLRNVALPFWK
GANGSGAFYGPYDPIAOTIYGHQSTEASAFDMSAMQVADNNLTDAIDAYOYA
KLAPEDSNAMVLAHNNQARYVEVASHVRFYKGGCKEMKSMENIHTIRINEEGCH
WSTIAKLTQVNFIMETLALMDLHVPKRCILITGPNTGKSAPFMSLTHVLG
RVLSVNSKSQFWLQPMSECKIALIDVTDPCMIYMDTYLRNGLDGIVSLDKHAP
MOTKFPALLTNSININNEVNYRLHRIKGFEPNFPKADNTEFFELTDQMSKF
ETRLMNOLELSDGEGENGESQSRFQCSARSANHEL"
2751. .4262
/gene="E2"
2751. .4262
/gene="E2"
/note="putative"
/codon_start=1
/product="regulatory protein E2"
/protein_id="AA079397.1"
/db_xref="GI:1020174"
/translation="MENLSDRFNVYLQDLNNIYESAANTIESQIEHWTLRKRAVLV
FARQGYTRLGYOYVPLAVSESRKQAIMQIOLQSEYAKEPMSLVDPISAELE
RSPENHFKKGPVSEVYIYNDKDNANAYIMMRYVYVDDDDQHKSPSGVNHGLEY
MOGTFHHYVLPADDAVSRYRTGHEWENVKRFPFAPVTSSTPDSGCGADNTSST
TPATTTSTSLSTSRKSOOTNTKGRVRRPSRTTRTQTHORRSRSKRSR
SRRLRSRSRSRSRSRSRSOSSDPOYRFRSGOVSILTTATTTTATNTSTRS
GRGSSSTSSSTSRPRPRRGALIGSSGRGRSSSTSPSKRSRGSESVRONGISP
DDVGKLSQSVSTRNTGRGLDELADLPVILVGEPTLCKFRNRAKLYAGLYKAF
STAMSWAGDGETRLGRSRLISFFSEQRKDFDKTVKYPKGVRSYGSFDSL"
3037. .4017
/gene="E4"
3037. .4017
/gene="E4"
/note="putative"

/codon_start=1
/protein_id="AA079398.1"
/db_xref="GI:1020175"
/translation="IPVORHLFALIKITSKKGGCOLRUFMTITKTMMLPCGDMFT
WMHTTNGIKVAVSTTQAYTICKELDTTMTYLLMVOYDAEIDIGTLIRKICLL
SPAPPHQHBQKQOTPPPPPPPLTPRDSRPPEVSHKPPKKEGGDGPV
GERPKRIGDGRGSPSGRGRGSDPDGPIPGDLNLTSTNNDSDPEGK
PSLPPPPPPPTTPEGGEGHPPPPPGNGHDEGPLEGAVGGDGHPPP
PAPMGHESSOLGNVASLTTTMSLFNQLVQELQVLDYWTKLSPQ"
4351. .5913
/gene="L2"
4351. .5913
/gene="L2"
/note="putative"
/codon_start=1
/product="minor capsid protein L2"
/protein_id="AA079399.1"
/db_xref="GI:1020176"
/translation="MARAKRYKRSATNIYRTCKOAGTCCPPDVINKVESTTADKILQ
YSGACVFFGGLISTGKGTGTVPLGEGDAVAVGNAPVIRPALVPTIGSDI
PVDLTNPVEPTTSITVPLDSTGPDLLGEEVETIAELHPTPPPPAVTTSNGSS
AVLEVAPEPTPSRVVRMYRTOYHNPSRQVITESPPTGESSLADHILVTSGGOTIG
GSTPELIELQDPKRSYSEIEPTTPRRTSTPIORIONIRRGSGULNRLOQVY
ENPLFVSRRSLVQFDPNPAEEVETQIFQDIDITFNEPPDRDLITIGRQYSE
TPAGVRSYRLGKRGITRTSGTQIGSOVHFYRLDSTINTEDPILDLGHSQDATI
VOGVESTFIIDINDENPLEDPSAHSDDLLEDENEDFSGSLVVGRRSTSYTVP
RETRRSQSYVQDTKGYVAVPEDRDTSDIIVTLPDPVYIHTPDTSGDFYLHPS
LSRKKRRKKY"
5929. .7485
/gene="L1"
5929. .7485
/gene="L1"
/note="putative"
/codon_start=1
/product="major capsid protein L1"
/protein_id="AA079400.1"
/db_xref="GI:1020177"
/translation="MAVMOAASGKYLPPSPVPARVOSTEDYVQRTNIYHAYSDRL
TVGHYFNYVDVNSAKIKYPKVSGNQHVRFLKLPDRFALADSVYNPDKERLYMA
CGITELGRQPLGVSGVHPLFNKYGDNENSSYTOPNSDRONVSEFDKLOMFI
TCGAPFLGEMHKALPCATDNPPOGSCPIELINRAIDGCMADIGYNLNRKALQON
RSDVSLDIVNETCKYPDLEKMONDVYGSCEFYARBOCYARHFFVRGKGTGDI
QIDEGSMKAAVYIPMNDQAOYKIGNSMFPVSSSLVSSDAQLENRPFMLORAGHN
NIGCFNOLFTVYVNTDRTNFTNSINWPNADVSKILENYKASEFOEYLRHVEVELSL
TLQCKVPLTAFLVAQIINAMNANLLEEMQLGVPAPDNDIHDYRYIDSAATRCDDKN
PKERBDPYKNNKFMVDVLTERTLSLDIOYSLGRKFLQAGLQDTTNGCTLSSRVS
TGIRKRRKN"

BASE COUNT 2426 a 1518 c 1680 g 2155 t
ORIGIN
Query Match 60.0%: Score 15; DB 14; Length 7779;
Best Local Similarity 100.0%: Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 941 ATCTGCTGTTTGAAG 955
Oy 3 atcgcgtgttgaag 17
|||||
Db 941 ATCTGCTGTTTGAAG 955

Search completed: December 26, 2001, 12:51:20
Job time: 8802 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:16:34 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title:	US-09-396-196F-4
Perfect score:	25
Sequence:	1 gqatctgctgttgaagcagcag 25

Scoring table: OLI60_NOC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

word size :

Total number of hits satisfying chosen parameters: 50995

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

post-processing: Listing first 100 summaries

Database :

1:	/SID2/gcgdata/genseq/genseqn/NA1980.DAT *
2:	/SID52/gcgdata/genseq/genseqn/NA1982.DAT *
3:	/SID52/gcgdata/genseq/genseqn/NA1982.DAT *
4:	/SID52/gcgdata/genseq/genseqn/NA1983.DAT *
5:	/SID52/gcgdata/genseq/genseqn/NA1984.DAT *
6:	/SID52/gcgdata/genseq/genseqn/NA1985.DAT *
7:	/SID52/gcgdata/genseq/genseqn/NA1986.DAT *
8:	/SID52/gcgdata/genseq/genseqn/NA1987.DAT *
9:	/SID52/gcgdata/genseq/genseqn/NA1988.DAT *
10:	/SID52/gcgdata/genseq/genseqn/NA1989.DAT *
11:	/SID52/gcgdata/genseq/genseqn/NA1990.DAT *
12:	/SID52/gcgdata/genseq/genseqn/NA1991.DAT *
13:	/SID52/gcgdata/genseq/genseqn/NA1992.DAT *
14:	/SID52/gcgdata/genseq/genseqn/NA1993.DAT *
15:	/SID52/gcgdata/genseq/genseqn/NA1994.DAT *
16:	/SID52/gcgdata/genseq/genseqn/NA1995.DAT *
17:	/SID52/gcgdata/genseq/genseqn/NA1996.DAT *
18:	/SID52/gcgdata/genseq/genseqn/NA1997.DAT *
19:	/SID52/gcgdata/genseq/genseqn/NA1998.DAT *
20:	/SID52/gcgdata/genseq/genseqn/NA1999.DAT *
21:	/SID52/gcgdata/genseq/genseqn/NA2000.DAT *
22:	/SID52/gcgdata/genseq/genseqn/NA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description
		% Match	Length	DB		
1	25	100.0	839	22	AA162941	Human genomic DNA
2	25	100.0	1041	20	AAAX01303	E. coli biotin synth.
3	25	100.0	1084	20	AAAN91329	E. coli Bio B gene.
4	25	100.0	1121	7	AAAN60486	Sequence encoding
5	25	100.0	5872	15	AAO62386	Biotin-biosynthesi
6	17	68.0	5504	20	AAK20559	Polynucleotide seq
7	16	64.0	801	22	AAH65216	C glutamylc acid codin
8	16	64.0	924	22	AAE71927	Corynebacterium gl
9	16	64.0	349980	22	AAH64966	C glutamic acid codin
10	15	60.0	810	22	AAH01365	Serratia marcescen
11	15	60.0	1869	20	AAZ23024	Rat Kd312 polypept

C	12	15	60.0	2022	21	AAC44133
C	13	15	60.0	3079	20	AAF623025
C	14	14	56.0	336	22	AAF65487
C	15	14	56.0	379	20	AAH87319
C	16	14	56.0	393	22	AAH10789
C	17	14	56.0	469	22	AAH12900
C	18	14	56.0	466	22	AAI34258
C	19	14	56.0	469	22	AAI02819
C	20	14	56.0	476	22	AAI22139
C	21	14	56.0	476	22	AAI47435
C	22	14	56.0	476	22	AAI07839
C	23	14	56.0	619	21	AAA44124
C	24	14	56.0	753	20	AAE00347
C	25	14	56.0	753	22	AAE89005
C	26	14	56.0	1173	20	AAE00358
C	27	14	56.0	1173	22	AAE89020
C	28	14	56.0	1244	21	AAE50896
C	29	14	56.0	1536	11	AAO04684
C	30	14	56.0	1699	22	AAI58178
C	31	14	56.0	1998	20	AAE00367
C	32	14	56.0	1998	22	AAE89029
C	33	14	56.0	2087	20	AAE00343
C	34	14	56.0	2087	22	AAE89001
C	35	14	56.0	2166	20	AAE00379
C	36	14	56.0	2166	22	AAE89041
C	37	14	56.0	2314	22	AAH17631
C	38	14	56.0	2319	22	AAH19233
C	39	14	56.0	2339	22	AAH07655
C	40	14	56.0	2402	22	AAE79014
C	41	14	56.0	2405	22	AAE27417
C	42	14	56.0	2447	22	AAH19186
C	43	14	56.0	2574	20	AAZ33977
C	44	14	56.0	2574	21	AAE37841
C	45	14	56.0	2574	21	AAE58239
C	46	14	56.0	2628	22	AAE66254
C	47	14	56.0	2669	20	AAE89020
C	48	14	56.0	2675	21	AAE98021
C	49	14	56.0	2845	22	AAH18012
C	50	14	56.0	7778	18	AAO44359
C	51	14	56.0	12284	11	AAO60061
C	52	14	56.0	12284	15	AAO71308
C	53	14	56.0	12284	17	AAE29521
C	54	14	56.0	12284	22	AAE30678
C	55	14	56.0	12563	22	AAE28546
C	56	14	56.0	124884	22	AAH74201
C	57	14	56.0	125157	22	AAH474202
C	58	14	56.0	349980	22	AAE88527
C	59	14	56.0	349980	22	AAE86528
C	60	13	52.0	17	16	AAO94366
C	61	13	52.0	17	19	AAV62508
C	62	13	52.0	23	16	AAO94365
C	63	13	52.0	23	19	AAV62509
C	64	13	52.0	80	22	AAI50377
C	65	13	52.0	128	21	AAE82189
C	66	13	52.0	157	21	AAE12133
C	67	13	52.0	192	21	AAE11863
C	68	13	52.0	251	19	AAI57288
C	69	13	52.0	263	22	AAI57288
C	70	13	52.0	299	21	AAE09435
C	71	13	52.0	300	21	AAA01377
C	72	13	52.0	300	22	AAE65466
C	73	13	52.0	367	22	AAI26911
C	74	13	52.0	367	22	AAI57131
C	75	13	52.0	384	18	AAV01888
C	76	13	52.0	391	22	AAE65133
C	77	13	52.0	398	22	AAE66455
C	78	13	52.0	400	21	AAE58844
C	79	13	52.0	408	22	AAE10988
C	80	13	52.0	423	20	AAE29977
C	81	13	52.0	437	20	AAE40711
C	82	13	52.0	450	22	AAI11056
C	83	13	52.0	450	22	AAI32311
C	84	13	52.0	450	22	AAI00977

Zee mares DNA fragm
 Rat kd312 genomic
 Novel human polyonu
 EST clone BP797
 Human cDNA clone (
 Probe #2833 for ge
 Probe #2944 used t
 Probe #2810 used f
 Probe #12072 for g
 Probe #16121 used
 Probe #7830 used t
 Human secreted exp
 Nucleotide sequenc
 Human FATP3 coding
 Nucleotide sequenc
 Murine FATP3 parti
 Nucleotide sequenc
 Murine FATP3 codin
 Nucleotide sequenc
 Human FATP3 coding
 Human cDNA sequenc
 Human secreted pro
 Human secreted pro
 Human FATP3 coding
 Human fatty acid t
 Human secreted pro
 Human PRO703 nucle
 Human PRO703 nucle
 Human PRO703 nucle
 C glutamicum codin
 Human H-beta 58 fa
 Human colon cancer
 Human cDNA sequenc
 Staphylococcus aur
 Sequence encoding
 Sequence encoding
 Hog cholera virus.
 Hog cholera virus
 Genomic fragment #
 Nucleotide sequenc
 Nucleotide sequenc
 C glutamicum codin
 C glutamicum codin
 Septoria nodorum I
 Septoria nodorum I
 Septoria nodorum s
 Septoria nodorum I
 Septoria nodorum s
 Probe #19058 used
 N. meningitidis pa
 Human secreted pro
 Human secreted pro
 Human secreted pro
 Human biallelic po
 Probe #25974 used
 Fusarium venenatum
 Human colon cancer
 C glutamicum codin
 Probe #16848 for g
 Probe #24401 used
 Human OTK27 gene.
 Novel human polyinu
 Novel human polynu
 Mycoplasma hyopne
 Cone snail cDNA en
 Fragment of human
 Human secreted pro
 Probe #987 for gene
 Probe #1003 used t
 Probe #905 used t

PA (NOVS) NOVARTIS FINANCE CORP.
 XX Patton DA;
 XX WPI: 1999-152902/13.
 DR P-PSDB; AAW73906.
 XX Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di:amino-pelargonic acid amino-transferase or biotin
 PT synthase
 XX Example 2: Column 37-40; 34pp: English.
 XX This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;
 Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 4.6e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ggatcgcgtgttgaagcgcagcag 25
 Db 63 ggatcgcgtgttgaagcgcagcag 87
 RESULT 3
 AAN91329
 ID AAN91329 standard: DNA: 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KW E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 FT
 PN GB216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 89GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee KA.
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp: English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in *S.cerevisiae*
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC *Lactobacillus*. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;
 Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 4.6e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ggatcgcgtgttgaagcgcagcag 25
 Db 86 ggatcgcgtgttgaagcgcagcag 110
 RESULT 4
 AAN60496
 ID AAN60496 standard: DNA: 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KW Biotin synthetase enzyme; E.coli; desbiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag=a
 FT
 PN JPE1149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure: Page 534; 23pp: Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desbiobiotin.
 XX
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;
 Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 4.6e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ggatcgcgtgttgaagcgcagcag 25
 Db 104 ggatcgcgtgttgaagcgcagcag 128
 RESULT 5
 AAO62386
 ID AAO62386 standard: DNA: 5872 BP.
 XX


```

XX 05-MAY-1999 (first entry)
DT Polynucleotide sequence from the genome of Treponema pallidum.
XX
DE Treponema pallidum infection: syphilis; Borrelia infection; animal;
KW enzyme production; ds.
XX
OS Treponema pallidum.
XX
PN W09859034-A2.
XX
PD 30-DEC-1998.
XX
PF 23-JUN-1998; 98WO-US13041.
XX
PR 24-JUN-1997; 97US-0050667.
XX
PA (HUMAN-) HUMAN GENOME SCI INC.
XX
PI Fraser CM;
XX
DR WPI; 1999-081273/07.
XX
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
XX
PS Claim 1; Page 519-522; 1150pp; English.
XX
CC AAX20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
SO Sequence 5504 BP; 1474 A; 1764 C; 1181 G; 1076 T; 9 other:

Query Match 68.0%; Score 17; DB 20; Length 5504:
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gctgttgaacgcagc 23
DB 1365 cctgtttgaagcgagc 1349

RESULT 7
AAH65216/c
ID AAH65216 standard; DNA; 801 BP.
XX
AC AAH65216;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 251.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.

```

```

XX (KYO) KYOMA HAKKO KOSYO KK.
FA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
PI WPI; 2001-376931/40.
XX P-PSDB: AAG89997.
DR
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
PS Claim 8; SEQ ID NO: 251; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from corynebacterium, and identifying a homologue of a gene derived
CC from corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SO Sequence 801 BP; 184 A; 221 C; 215 G; 181 T; 0 other:

Query Match 64.0%; Score 16; DB 22; Length 801:
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ctagttgaacgcagc 23
DB 784 CTGTTGAAGCGCAGC 769

RESULT 8
AAAF71927/c
ID AAFA71927 standard; DNA; 924 BP.
XX
AC AAFA71927;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO: 349.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; dtol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme; ds.
XX
OS Corynebacterium glutamicum.
XX
PN W0200100843-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-1800923.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.

```


OY 8 cgtgtgaagcgagc 23
|
Db 238128 CTGTTGACGCAGC 238113

RESULT 10
AAH01365
ID AAH01365 standard; DNA: 810 BP.
XX
AC AAH01365;
XX
DT 24-JUL-2001 (first entry)
XX

Serratia marcescens nucleotide sequence SEQ ID NO:1356.

Species specific; genus specific; family specific; probe; detection;
KM identification; algal; archaeal; bacterial; fungal; parasitica;
KM microorganism; diagnosis; translation elongation factor Tu; toxin;
KM translation elongation factor G; RecA recombinase; resistance;
KM catalytic subunit of proton-translocating ATPase; antimicrobial;
KM vaccine; primer; ds.
XX
OS Serratia marcescens.
XX
PN W0200123604-A2.
XX
PD 05-APR-2001.
XX
PE 28-SEP-2000; 2000MO-CA01150.
XX
PF 28-SEP-1999; 99CA-2283458.
XX
PR 19-MAY-2000; 2000CA-2307010.
XX

(INFE-) INFECTIO DIAGNOSTIC (TDI) INC.

PA Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
PI Picard FJ, Roy PH;
XX

WPI: 2001-245006/25.

Nucleic acid sequences are used to generate universal probes and
primers which can be used to identify and detect the presence of algal,
archaeal, bacterial, fungal and parasitica species in a test sample -
XX
PS Disclosure: Page 1131-1132; 1580pp; English.

The present invention describes a method for generating a repository of
nucleic acids of tuf, fus, apd and/or recA genes from which probes
and/or primers are derived. The method comprises amplifying the nucleic
acids of determined algal, archaeal, bacterial, fungal and parasitica
species with a combination of defined primer pairs. The method can be
used for producing probes and/or primers for detecting one or more
related microorganisms e.g. algae, archaea, bacteria, fungi and
parasites, for universal detection and for specific and ubiquitous
detection and identification of an algal, archaeal, bacterial, fungal
and parasitica species, genus, family and group. A nucleic acid (II)
obtained using the method of the invention can be used for the universal
detection of any bacterium, fungus or parasite in a sample and for the
detection of at least one antimicrobial agent resistance gene or at
least one toxin gene. hexa nucleic acids are used for the specific and
ubiquitous detection and for identification of Streptococcus pneumoniae.
(I) can be used to design a therapeutic agent which is effective against
microorganisms. Microbial species or genus or family or phylum or group
which can be detected include Abiotrophia adiacens, Bordetella sp.,
Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
Mycobacteriaceae family, Pseudomonas group, Streptococcus sp.,
Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
provides faster results than substrate specificity tests as results can
be determined in an hour and improved accuracy is also achieved.
XX
CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
CC which are given in the exemplification of the present invention.
XX
SQ Sequence 810 BP; 129 A; 270 C; 268 G; 143 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 810;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 cgtgtgaagcgagc 22
|
Db 742 cgtgtgaagcgagc 756

RESULT 11
AAZ23024/C
ID AAZ23024 standard; cDNA: 1689 BP.
XX
AC AAZ23024;
XX
DT 17-JAN-2000 (first entry)
XX

Rat kd312 polypeptide encoding cDNA.

XX kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW heart attack; head trauma; neurodegenerative disease; rat;
KW Parkinson's disease; Alzheimer's disease; SS.
XX

OS Rattus sp.

PN W09950288-A2.

PD 07-OCT-1999.

PF 30-MAR-1999; 99WO-US06993.

PR 31-MAR-1998; 98US-0053374.

PA (AMGE-) AMGEN INC.

PI Yen K;
XX

WPI: 1999-601322/51.

DR P-PSDB; AAY42694.

kd312 polypeptides useful for treating diseases and disorders
PT associated with alterations in cell proliferation and cell death -
XX
PS Claim 2; Fig 9; 85pp; English.

The invention provides nucleic acid molecules encoding human and rat
kd312 polypeptides. The kd312 polypeptides can be expressed by standard
recombinant methodology. The kd312 sequences, and the antibodies against
the proteins may be used to treat or diagnose the presence or progression
of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
Parkinson's disease and Alzheimer's disease). The present sequence
represents the rat kd312 cDNA sequence.
XX

Sequence 1689 BP; 417 A; 462 C; 434 G; 376 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 1689;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggatcgtgttga 15
|
Db 495 GGATCTGCTGTTGA 481

RESULT 12
AAC44132/C
ID AAC44132 standard; DNA: 2022 BP.
XX
AC AAC44132;

```
XX 18-OCT-2000 (first entry)
DT
XX
DE Zea mays DNA fragment SEQ ID NO: 41750.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-014732.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
```

```
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 60.0%; Score 15; DB 21; Length 2022;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4 tctgctgtttgaagc 18
Db 1268 TCTGCTGTTGAAGC 1254
RESULT 13
AA223025/C
ID AA223025 standard; DNA; 3079 BP.
AC AA223025;
XX
XX 17-JAN-2000 (first entry)
XX
XX Rat kd312 genomic DNA sequence.
XX
XX kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
XX heart attack; head trauma; neurodegenerative disease; rat;
KW
```

```
KW Parkinson's disease; Alzheimer's disease; ss.
XX
XX Rattus sp.
OS
XX W09950288-A2.
PN
XX
XX 07-OCT-1999.
PD
XX
XX 30-MAR-1999; 99WO-US06993.
PF
XX
XX 31-MAR-1998; 98US-0053374.
PR
XX
XX (AMGE-) AMGEN INC.
XX
XX Yen K;
PI
XX WPI: 1999-601322/51.
DR
XX P-PSDB; AAY42694.
PT
XX kd312 polypeptides useful for treating diseases and disorders
XX associated with alterations in cell proliferation and cell death
XX
XX Claim 2: Fig 7: 85pp: English.
PS
XX
XX The invention provides nucleic acid molecules encoding human and rat
XX kd312 polypeptides. The kd312 polypeptides can be expressed by standard
XX recombinant methodology. The kd312 sequences, and the antibodies against
XX the proteins may be used to treat or diagnose the presence or progression
XX of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
XX stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
XX Parkinson's disease and Alzheimer's disease). The present sequence
XX represents the rat kd312 genomic DNA sequence.
SQ
XX Sequence 3079 BP; 708 A; 843 C; 829 G; 699 T; 0 other;
```

```
Query Match 60.0%; Score 15; DB 20; Length 3079;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ggatcgtcgtttga 15
Db 803 GGATCTGCTTTGA 789
RESULT 14
AAF65487/C
ID AAF65487 standard; cDNA; 336 BP.
AC AAF65487;
XX
XX 09-APR-2001 (first entry)
XX
XX Novel human polynucleotide, SEQ ID NO: 1243.
DE
XX Human: cytostatic; gene therapy; colon cancer; prostate cancer;
XX breast cancer; lung cancer; cancer detection; ss.
KW
XX Homo sapiens.
OS
XX W0200102568-A2.
XX
XX 11-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18374.
XX
XX 02-JUL-1999; 99US-0142310.
XX
XX 02-JUL-1999; 99US-0142311.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
```


Query Match 56.0%; Score 14; DB 22; Length 469;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgtttga 15
 |||||
 Db 295 gatctgctgtttga 308

RESULT 19

AAI02819
 ID AAI02819 standard; DNA; 469 BP.

XX
 AC AAI02819;

XX DT 09-OCT-2001 (first entry)

DE Probe #2810 used to measure gene expression in human breast sample.

XX KW Probe: human; breast disease; breast cancer; development disorder; ss;
 inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX OS Homo sapiens.

XX PN WO200157270-A2.

XX PD 09-AUG-2001.

XX PF 29-JAN-2001; 2001WO-US00661.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression

XX in a human breast -

XX Claim 25; SEQ ID No 2810; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes.

XX The present sequence is one such probe. The probes are useful for

XX measuring human gene expression in a human breast sample, where the probe

XX hybridises at high stringency to a nucleic acid expressed in the human

XX breast. The probes are useful for predicting, diagnosing, grading,

XX straging, monitoring and prognosing diseases of the human breast,

XX particularly those diseases with polygenic aetiology. The diseases

XX include: breast cancer, disorders of development, inflammatory diseases

XX of the breast, fibrocystic changes, proliferative breast disease and

XX non-carcinoma tumours.

RESULT 20

AAI22139
 ID AAI22139 standard; DNA; 476 BP.

XX AC AAI22139;

XX DT 12-OCT-2001 (first entry)

DE Probe #12072 for gene expression analysis in human cervical cell sample.

XX KW Probe: human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID No 12072; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SEMP). The present sequence is one such probe. The SEMPs are derived

XX from human HeLa cells. The SEMPs can be used to produce a single exon

XX microarray, which can be used for measuring human gene expression in a

XX sample derived from human cervical epithelial cells. By measuring gene

XX expression, the probes are therefore useful in grading and/or staging

XX of diseases of the cervix, notably cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;

XX Query Match 56.0%; Score 14; DB 22; Length 476;

XX Best Local Similarity 100.0%; Pred. No. 63;

XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgtttga 15
 |||||

Db 135 gatctgctgtttga 148

RESULT 21

AAI47435
 ID AAI47435 standard; DNA; 476 BP.

XX AC AAI47435;

XX DT 17-OCT-2001 (first entry)

DE Probe #16121 used to measure gene expression in human placenta sample.

XX Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX

OS Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-488897/53.

DR Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

PT Claim 25; SEQ ID No 16121; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders.

XX Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;

SO Query Match 56.0%; Score 14; DB 22; Length 476;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatctgctgtttga 15

DB 135 gatctgctgtttga 148

RESULT 22

AA107839

ID AA107839 standard; DNA; 476 BP.

XX AA107839;

XX 09-OCT-2001 (first entry)

DE Probe #7830 used to measure gene expression in human breast sample.

XX Probe: human; breast disease; breast cancer; development disorder; ss;

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX WO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-476286/51.

DR Novel single exon nucleic acid probe used to measuring gene expression

XX in a human breast -

PT Claim 25; SEQ ID No 7830; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes.

CC The present sequence is one such probe. The probes are useful for

CC measuring human gene expression in a human breast sample, where the probe

CC hybridises at high stringency to a nucleic acid expressed in the human

CC breast. The probes are useful for predicting, diagnosing, grading,

CC staging, monitoring and prognosing diseases of the human breast.

CC particularly those diseases with polygenic aetiology. The diseases

CC include: breast cancer, disorders of development, inflammatory diseases

CC of the breast, fibrocystic changes, proliferative breast disease and

CC non-carcinoma tumours.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;

SO Query Match 56.0%; Score 14; DB 22; Length 476;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatctgctgtttga 15

DB 135 gatctgctgtttga 148

RESULT 23

AAA44124/c

ID AAA44124 standard; cDNA; 619 BP.

XX AAA44124;

XX 21-AUG-2000 (first entry)

DE Human secreted expressed sequence tag SEQ ID NO:699.

XX Human: mouse; chicken; rat; secreted expressed sequence tag; sEST;

KW expressed sequence tag; EST; probe; chemokine; proliferative;

KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;

KW thrombolytic; antiinflammatory; cytoskeletal; antibacterial; antifungal;

KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;

KW anticancer; osteoprotective; neuroprotective; neotropic; antipsoriatic;

KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;

KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;

KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;

KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;

KW central nervous system disorder; Alzheimer's disease; stroke;

KW Parkinson's disease; Huntington's disease; coagulation disorder;

KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;

KW tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

XX WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24206.
 XX 15-OCT-1998; 98US-0104436.
 XX (GENY) GENETICS INST INC.
 XX Jacobs K, McCoy JM, Lavallic ER, Collins-Racie LA, Evans C,
 PI Merberg D, Treacy M, Bowman MR;
 XX WPI: 2000-317938/27.
 DR Isolated polynucleotides, and encoded proteins, comprising secreted
 XX expressed sequence tags (ESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 XX
 XX Claim 1: Page 383; 803pp: English.
 XX AAA3426 to AAA5925 represent specifically claimed secreted expressed
 CC sequence tags (ESTs), isolated from human, mouse, chicken and rat
 CC tissue sources. The ESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytoskeletal; antibacterial; antifungal; antiviral; antidiabetic;
 CC antiasthmatic; vulnery; antitumor; osteoprotic; neuroprotective;
 CC neurotrophic; antiparkinsonian; antipsoriatic; cerebroprotective;
 CC anticonvulsant; and antidepressant. The ESTs can be used for gene
 CC therapy and in vaccines. The ESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the ESTs. Proteins encoded by the ESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA5926 to AAA5931 represent linker variants which are given
 CC in the exemplification of the present invention.
 XX
 SQ Sequence 619 BP; 193 A; 123 C; 125 G; 178 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 619;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttga 16
 |||||
 Db 43 ATCTGCTGTTGA 30

RESULT 24
 ID AA200347C
 ID AA200347 standard: DNA; 753 BP.
 AC AA200347;
 XX
 XX 26-OCT-1999 (first entry)
 DT
 XX
 DE Nucleotide sequence of human hsfATP3.
 XX
 XX Fatty acid transport protein: FATP; long chain fatty acid; LCFA; human;
 KW fatty acid: FATP biosynthesis; obesity; diabetes; heart disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO936537-A2.
 XX
 PD 22-JUL-1999.
 XX

PE 14-JAN-1999; 99WO-US00182.
 XX
 XX 14-JAN-1999; 99US-0232201.
 PR 15-JAN-1998; 98US-0071374.
 PR 20-JUL-1998; 98US-0093491.
 PR 04-DEC-1998; 98US-0110941.
 PR 14-JAN-1999; 99US-0232195.
 PR 14-JAN-1999; 99US-0232197.
 PR 14-JAN-1999; 99US-0232200.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 XX
 PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
 XX
 XX WPI: 1999-444398/37.
 DR P-PSDB; AAY14937.
 XX
 PT Fatty acid transport proteins and related polynucleotides, useful
 PT for treating obesity, diabetes and heart disease
 XX
 XX Example 1: Fig 16; 255pp: English.
 XX
 CC The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating synthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.
 XX
 SQ Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 753;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttga 17
 |||||
 Db 436 TCTGCTGTTGA 423

RESULT 25
 ID AAF89005C
 ID AAF89005 standard: DNA; 753 BP.
 AC AAF89005;
 XX
 XX 06-JUL-2001 (first entry)
 DT
 XX
 DE Human FATP3 coding sequence SEQ ID NO: 14.
 XX
 XX Fatty acid transport protein: FATP; human; mouse; rat; rice blast fungus;
 KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidemia;
 KW weight control; tuberculosis; TB; anti-fungal; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200121795-A2.
 XX
 PD 29-MAR-2001.
 XX

PR 21-SEP-2000; 2000MO-US25891.
 XX 23-SEP-1999; 99US-0405504.
 PR 23-SEP-1999; 99US-0405505.
 PR 16-DEC-1999; 99US-0465280.
 PR 17-FEB-2000; 2000US-0506252.
 PR 06-JUL-2000; 2000US-0611197.
 XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 XX (MILL-) MILLENNIUM PHARM INC.
 PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
 XX WPI; 2001-354783/37.
 DR P-PSDB; AAB83227.
 XX
 PT New fatty acid transport proteins (FATPs) useful for the manufacture of
 PT medicament for treating obesity, diabetes and heart disease -
 PS Disclosure; Fig 16; 287pp; English.
 XX
 CC The present invention provides the protein and coding sequences of fatty
 CC acid transport proteins (FATPs) from a number of species, including
 CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
 CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus
 CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
 CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
 CC tuberculosis can be used to identify inhibitors which can then be used to
 CC treat TB. That from M. grisea (also known as rice blast fungus) can be
 CC used to develop anti-fungal agents capable of preventing infection for
 CC rice. Those from the human can be used to develop treatments for
 CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
 CC present sequence is one of the sequences described in the exemplification
 CC of the invention.
 CC
 SQ Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 753;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||||
 Db 436 tctgctgttgaag 423

RESULT 26
 AAZ00358/C
 ID AAZ00358 standard; DNA; 1173 BP.
 XX
 AC AAZ00358;
 XX
 DT 26-OCT-1999 (first entry)
 XX
 DE Nucleotide sequence of human hsfFATP3.
 XX
 KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9936537-A2.
 XX
 PD 22-JUL-1999.
 XX
 PF 14-JAN-1999; 99WO-US00182.
 XX
 PR 14-JAN-1999; 99US-0232201.
 PR 15-JAN-1999; 99US-0071374.
 PR 20-JUL-1999; 98US-0093491.
 PR 04-DEC-1999; 98US-0110941.
 PR 14-JAN-1999; 99US-0232195.

PR 14-JAN-1999; 99US-0232197.
 PR 14-JAN-1999; 99US-0232200.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX (WHED) WHITEHEAD INST BIOMEDICAL RES.
 XX
 PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
 XX WPI; 1999-444398/37.
 DR P-PSDB; AAV14948.
 XX
 PT Fatty acid transport proteins and related polynucleotides, useful
 PT for treating obesity, diabetes and heart disease
 PS Claim 31; Fig 48; 255pp; English.
 XX
 CC The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.
 CC
 SQ Sequence 1173 BP; 247 A; 327 C; 351 G; 248 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1173;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||||
 Db 853 tctgctgttgaag 840

RESULT 27
 AAF89020/C
 ID AAF89020 standard; DNA; 1173 BP.
 XX
 AC AAF89020;
 XX
 DT 06-JUL-2001 (first entry)
 XX
 DE Human FATP3 partial coding sequence SEQ ID NO: 50.
 XX
 KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
 KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
 KW weight control; tuberculosis; TB; anti-fungal; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200121795-A2.
 XX
 PD 29-MAR-2001.
 XX
 PF 21-SEP-2000; 2000MO-US25891.
 XX
 PR 23-SEP-1999; 99US-0405504.
 PR 23-SEP-1999; 99US-0405505.
 PR 16-DEC-1999; 99US-0465280.
 PR 17-FEB-2000; 2000US-0506252.
 PR 06-JUL-2000; 2000US-0611197.

XX (MHED) WHITEHEAD INST. BIOMEDICAL RES.
 PA (MILL.) MILLENNIUM PHARM INC.
 XX
 PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
 DR WPI: 2001-354783/37.
 DR P-PSDB: AAB83248.
 XX
 PT New fatty acid transport proteins (FATPs) useful for the manufacture of
 PT medication for treating obesity, diabetes and heart disease -
 XX
 PS Disclosure: Fig 48; 287pp; English.
 XX
 CC The present invention provides the protein and coding sequences of fatty
 CC acid transport proteins (FATPs) from a number of species, including
 CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
 CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus
 CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
 CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
 CC tuberculosis can be used to identify inhibitors which can then be used to
 CC treat TB. That from M. grisea (also known as rice blast fungus) can be
 CC used to develop anti-fungal agents capable of preventing infection of
 CC rice. Those from the human can be used to develop treatments for
 CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
 CC present sequence is one of the sequences described in the exemplification
 CC of the invention.
 CC
 SO Sequence 1173 BP; 247 A; 327 C; 351 G; 248 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1173;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctcgtgttgaag 17
 |||
 DB 853 TCTGCTGTTGAAG 840

RESULT 28
 ID AAA50896/c
 AC AAA50896 standard; cDNA; 1244 BP.
 XX
 AC AAA50896;
 XX
 DT 28-SEP-2000 (first entry)
 XX
 DE Soybean PI metabolism protein clone, ssm.pk0068.c5 partial cDNA.
 XX
 KW Enzyme: phosphatidylinositol-4-phosphate kinase; PI;
 KW phosphatidylinositol metabolism; signal transduction; membrane anchor;
 KW membrane trafficking pathway regulator; vacuole homeostasis regulator;
 KW transgenic plant; soybean; EST; expressed sequence tag; ss.
 XX
 OS Glycine max.
 XX
 FH Key Location/Qualifiers
 FT CDS 2..991
 FT /*tag= a
 FT /partial
 FT /product= "ssm.pk0068.c5"
 XX
 PN WO200036119-A2.
 XX
 PD 22-JUN-2000.
 XX
 PF 17-DEC-1999; 99WO-US30182.
 XX
 PR 18-DEC-1998; 98US-0112925.
 XX
 PA (DUPD) DU PONT DE NEMOURS & CO E I.
 XX

PI Allen SM, Kinney AJ, Miao G, Rafalski JA, Sakai H, Weng Z;
 XX
 DR WPI: 2000-431597/37.
 DR P-PSDB: AAY9751.
 XX
 PT New isolated polynucleotide encoding phosphatidylinositol metabolism
 PT protein is useful for producing transgenic plants with an altered level
 PT of phosphatidylinositol -
 XX
 PS Claim 2; Page 50-51; 65pp; English.
 XX
 CC Phosphatidylinositol, PI, may be modified in cells to have roles in
 CC signal transduction, membrane associated protein anchoring and regulation
 CC of membrane trafficking pathways. The enzyme phosphatidylinositol-4-
 CC phosphate 5-kinase, converts phosphatidylinositol-4-phosphate to
 CC phosphatidylinositol-4,5-diphosphate. Phosphatidylinositol-4,5-
 CC diphosphate is thought to play a role in vacuole homeostasis regulation.
 CC The present sequence is the partial coding sequence of the soybean PI
 CC metabolism protein clone, ssm.pk0068.c5, isolated from an expressed
 CC sequence tag, EST. The present sequence may be used to create transgenic
 CC plants which express phosphatidylinositol-4-phosphate kinase, and hence
 CC be useful for modifying plant phosphatidylinositols.
 CC
 SO Sequence 1244 BP; 392 A; 207 C; 274 G; 371 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1244;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gtttgaagcgacg 23
 |||
 DB 886 GTTTGAAGCGACG 873

RESULT 29
 ID AAQ04684
 AC AAQ04684 standard; DNA; 1536 BP.
 XX
 AC AAQ04684;
 XX
 DT 05-OCT-1990 (first entry)
 XX
 DE SalI/KpnI fragment of pUOI contg. haloacetate dehalogenase H-1.
 XX
 KW Haloacetate dehalogenase; H-1; H-2; ss.
 XX
 OS Moraxell sp. B.
 XX
 FH Key Location/Qualifiers
 FT CDS 413..1296
 FT /*tag= a
 FT /label=H-1_gene
 XX
 PN JP02109981-A.
 XX
 PD 23-APR-1990.
 XX
 PF 08-APR-1988; 88JP-0085393.
 XX
 PR 08-APR-1988; 88JP-0085393.
 XX
 PA (MITN) MITSUBISHI GAS CHEM KK.
 XX
 DR WPI: 1990-168356/22.
 DR P-PSDB: AAR06435.
 XX
 PF Haloacetate dehalogenase coding gene - used to raise
 PF decomposition activity of arganac halogenated cpds.
 XX
 PS Claim 1; Fig 2; 8pp; Japanese.
 XX
 CC The DNA can be ligated into a vector which can then be used to

CC transform E.coli for the prodn. of H-1. The protein is used in
CC the decomposition of organic halogenated cpds.
CC See also AA004685.

XX Sequence 1536 BP; 315 A; 445 C; 449 G; 326 T; 1 other;

Query Match 56.0%; Score 14; DB 11; Length 1536;

Best Local Similarity 100.0%; Pred. No. 64;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgcagcag 25
|||||
Db 320 ttgaagcgcagcag 333

RESULT 30

AA158178/c
ID AA158178 standard; cDNA; 1699 BP.

XX AA158178;

DE 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 381.

XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX peripheral nervous system; neuropathy; central nervous system; CNS;

XX Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;

XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX leukemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

PD 26-JUL-2001.

PE 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSEQ-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

DR P-PSDB: AAM39022.

XX Novel nucleic acids and polypeptides, useful for treating disorders

CC such as central nervous system injuries -

PS Claim 1; SEQ ID NO 381; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and

CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, Leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 1699 BP; 340 A; 490 C; 499 G; 370 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1699;

Best Local Similarity 100.0%; Pred. No. 64;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgcgtttgaag 17
|||||
Db 1393 tctgcgtttgaag 1380

RESULT 31

AAZ00367/c
ID AAZ00367 standard; DNA; 1998 BP.

XX AAZ00367;

DE 26-OCT-1999 (first entry)

XX Nucleotide sequence of murine mmFATP3.

XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;

XX fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.

XX Mus sp.

XX WO936537-A2.

PD 22-JUL-1999.

PE 14-JAN-1999; 99WO-US00182.

XX 14-JAN-1999; 99US-0232201.

PR 15-JAN-1998; 98US-0071374.

PR 20-JUL-1998; 98US-0093491.

PR 04-DEC-1998; 98US-0110941.

PR 14-JAN-1999; 99US-0232195.

PR 14-JAN-1999; 99US-0232197.

PR 14-JAN-1999; 99US-0232200.

XX (MILL-) MILLENNIUM PHARM INC.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

PI WPI: 1999-444398/37.

DR P-PSDB: AAY14957.

XX Fatty acid transport proteins and related polynucleotides, useful

PT for treating obesity, diabetes and heart disease

XX Example 1; Fig 66; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)

CC that mediate transport of long chain fatty acids (LCFAs) across cell

CC membranes into cells. Human and murine FATP proteins and nucleic acids

CC encoding the proteins are provided. The FATP proteins can be produced

CC by standard recombinant methodology. Fatty acid uptake by cells can be

CC modulated by modulating biosynthesis of FATP proteins especially FATP6.

CC In particular, antisense oligonucleotides can be used to modulate FATP

CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid

CC uptake in cardiac muscle of humans. Agents can be directed to cardiac

CC muscle or liver by administration of a complex of the agent and a FATP6

CC binding moiety. DNA encoding FATP proteins can be used as a reference

CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.

XX Sequence 1998 BP; 422 A; 567 C; 597 G; 412 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 20; Length 1998;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
 |||||

Db 1690 TCTGCTGTTCGAAG 1677

RESULT 32

AAAF89029/c
 ID AAF89029 standard; DNA: 1998 BP.

AC AAF89029;

DT 06-JUL-2001 (first entry)

DE Murine FATP3 partial coding sequence SEQ ID NO: 68.

KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
 KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
 KW weight control; tuberculosis; TB; anti-fungal; ds.

XX Mus musculus.

XX WO200121795-A2.

XX 29-MAR-2001.

XX 21-SEP-2000; 2000WO-US25891.

XX 23-SEP-1999; 99US-0405504.

XX 23-SEP-1999; 99US-0405505.

XX 16-DEC-1999; 99US-0465280.

XX 17-FEB-2000; 2000US-0506252.

XX 06-JUL-2000; 2000US-0611197.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PA (MILL-) MILLENNIUM PHARM INC.

PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;

XX WPI: 2001-354783/37.

XX P-PSDB; AAB83257.

XX New fatty acid transport proteins (FATPs) useful for the manufacture of

XX medicament for treating obesity, diabetes and heart disease -

XX Disclosure; Fig 66: 287pp; English.

XX The present invention provides the protein and coding sequences of fatty
 CC acid transport proteins (FATPs) from a number of species, including
 CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
 CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus
 CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
 CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
 CC tuberculosis can be used to identify inhibitors which can then be used to
 CC treat TB. That from M. grisea (also known as rice blast fungus) can be
 CC used to develop anti-fungal agents capable of preventing infection of
 CC rice. Those from the human can be used to develop treatments for
 CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
 CC present sequence is one of the sequences described in the exemplification
 CC of the invention.

XX
 SQ Sequence 1998 BP; 422 A; 567 C; 597 G; 412 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 22; Length 1998;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
 |||||

Db 1690 TCTGCTGTTCGAAG 1677

RESULT 33

AAZ00343/c
 ID AAZ00343 standard; DNA: 2087 BP.

XX AAZ00343;

DT 26-OCT-1999 (first entry)

DE Nucleotide sequence of murine mmFATP3.

KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.

XX Mus musculus.

XX WO9936537-A2.

XX 22-JUL-1999.

XX 14-JAN-1999; 99WO-US00182.

XX 14-JAN-1999; 99US-0232201.

XX 15-JAN-1998; 98US-0071374.

XX 20-JUL-1998; 98US-0093491.

XX 04-DEC-1998; 98US-0110941.

XX 14-JAN-1999; 99US-0232195.

XX 14-JAN-1999; 99US-0232197.

XX 14-JAN-1999; 99US-0232200.

PA (MILL-) MILLENNIUM PHARM INC.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

XX WPI: 1999-44398/37.

XX P-PSDB; AAV14933.

XX Fatty acid transport proteins and related polynucleotides, useful

XX for treating obesity, diabetes and heart disease

XX Example 1; Fig 8; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.

Sequence 2087 BP; 441 A; 596 C; 624 G; 426 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2087;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgtttgaag 17
|||||
Db 1779 TCTGCTGTTTGAAG 1766

RESULT 34

AAf89001/c
ID AAF89001 standard; DNA: 2087 BP.

XX AAF89001;

XX 06-JUL-2001 (first entry)

XX Murine FATP3 coding sequence SEQ ID NO: 6.

XX Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
KM weight control; tuberculosis; TB; anti-fungal; ds.

XX Mus musculus.

XX WO200121795-A2.

XX 29-MAR-2001.

XX 21-SEP-2000; 2000WO-US25891.

XX 23-SEP-1999; 99US-0405504.

XX 23-SEP-1999; 99US-0405505.

XX 16-DEC-1999; 99US-0465280.

XX 17-FEB-2000; 2000US-0506252.

XX 06-JUL-2000; 2000US-0611197.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX (MILL-) MILLENNIUM PHARM INC.

XX Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;

XX WPI; 2001-354783/37.

XX P-PSDB; AAB83223.

XX New fatty acid transport proteins (FATPs) useful for the manufacture of

XX medicament for treating obesity, diabetes and heart disease -

XX Examples; Fig 8; 287pp; English.

XX The present invention provides the protein and coding sequences of fatty
CC acid transport proteins (FATPs) from a number of species, including
CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus
CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
CC treat TB. That from M. grisea (also known as rice blast fungus) can be
CC used to develop anti-fungal agents capable of preventing infection of
CC rice. Those from the human can be used to develop treatments for
CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
CC present sequence is one of the sequences described in the exemplification
CC of the invention.

XX Sequence 2087 BP; 441 A; 596 C; 624 G; 426 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2087;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgtttgaag 17
|||||
Db 1779 TCTGCTGTTTGAAG 1766

RESULT 35

AAZ00379/c
ID AAZ00379 standard; DNA: 2166 BP.

XX AAZ00379;

XX 26-OCT-1999 (first entry)

XX Nucleotide sequence of human hFATP5 gene.

XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;

XX fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.

XX Homo sapiens.

XX WO9936537-A2.

XX 14-JAN-1999; 99WO-US00182.

XX 14-JAN-1999; 99US-0232201.

XX 15-JAN-1998; 98US-0071374.

XX 20-JUL-1998; 98US-0093491.

XX 04-DEC-1998; 98US-0110941.

XX 14-JAN-1999; 99US-0232195.

XX 14-JAN-1999; 99US-0232197.

XX (MTL-) MILLENNIUM PHARM INC.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

XX WPI; 1999-444398/37.

XX P-PSDB; AAT14969.

XX Fatty acid transport proteins and related polynucleotides, useful

XX for treating obesity, diabetes and heart disease

XX Claim 31; Fig 94A-B; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase circulating fatty acids. Blocking the
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.

XX Sequence 2166 BP; 370 A; 682 C; 695 G; 419 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2166;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgtttgaag 17
 |||
 DB 1987 TCTGCTGTTTGAAG 1974

RESULT 36
 AAF89041/C
 ID AAF89041 standard; DNA: 2166 BP.
 XX
 AC AAF89041;
 XX
 DT 06-JUL-2001 (first entry)
 XX
 DE Human FATP3 coding sequence SEQ ID NO: 101.
 XX
 KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
 KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
 KW weight control; tuberculosis; TB; anti-fungal; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200121795-A2.
 XX
 PD 29-MAR-2001.
 XX
 PF 21-SEP-2000; 2000MO-US25891.
 XX
 PR 23-SEP-1999; 99US-0405504.
 PR 23-SEP-1999; 99US-0405505.
 PR 16-DEC-1999; 99US-0465280.
 PR 17-FEB-2000; 2000US-0506252.
 PR 06-JUL-2000; 2000US-0611197.
 XX
 PA (WHEED) WHITEHEAD INST BIOMEDICAL RES.
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
 DR WPL; 2001-354783/37.
 DR P-PSDB; AAB83278.
 XX
 PT New fatty acid transport proteins (FATPs) useful for the manufacture of
 PT medicament for treating obesity, diabetes and heart disease -
 XX
 PS Disclosure; Fig 94; 287pp; English.

CC The present invention provides the protein and coding sequences of fatty
 CC acid transport proteins (FATPs) from a number of species, including
 CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
 CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus
 CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
 CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
 CC treat TB. That from M. grisea (also known as rice blast fungus) can be
 CC used to develop anti-fungal agents capable of preventing infection of
 CC rice. Those from the human can be used to develop treatments for
 CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
 CC present sequence is one of the sequences described in the exemplification
 CC of the invention.
 CC
 XX
 SO Sequence 2166 BP; 370 A; 682 C; 695 G; 419 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2166;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgtttgaag 17
 |||
 DB 1987 TCTGCTGTTTGAAG 1974

RESULT 37

AAH17621/C
 ID AAH17621 standard; cDNA: 2314 BP.
 XX
 AC AAH17621;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:17145.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
 XX
 DR WPL; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 17145; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC
 XX
 SO Sequence 2314 BP; 402 A; 704 C; 747 G; 461 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2314;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgtttgaag 17
 |||
 DB 2022 TCTGCTGTTTGAAG 2009

RESULT 38
AAH19223/c
ID AAH19223 standard; cDNA: 2319 BP.
XX
XX
AC AAH19223:
XX
XX 17-JUL-2001 (first entry)
XX
DE Human secreted protein-encoding gene 17 cDNA clone HSLCAL5, SEQ ID NO:64.
XX
XX Human: secreted protein; proliferative disorder; cancer;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX inflammation; allergy; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;
XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
XX cardiovascular disorder; angiotensin disorder; kidney disorder;
XX gastrointestinal disorder; pregnancy-related disorder; tumour;
XX endocrine disorder; infection; wound healing; vulnery;
XX cell culture; chemotaxis; food additive;
XX binding partner identification; ss.
XX
XX Homo sapiens.
XX
XX WO200132910-A2.
XX
XX 10-MAY-2001.
XX
XX 25-OCT-2000; 2000MO-US29362.
XX
XX 29-OCT-1999; 99US-0162240.
XX 30-JUN-2000; 2000US-0215131.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Birse CE, Ni J, Soppet DR:
XX
XX WPI: 2001-335835/35.
XX N-PSDB: AAG62209.
XX
XX Novel 27 isolated human secreted proteins and polynucleotides encoding
XX them useful for treating, diagnosing, preventing Alzheimer's disease,
XX Parkinson's disease, AIDS, rheumatoid arthritis, asthma
XX
XX Claim 4; Page 478-479; 594pp; English.
XX
XX AAH19170-AAH19250 represent cDNAs corresponding to 27 human secreted
XX protein genes, and AAG62156-AAG62235 represent the proteins they encode.
XX AAG62236-AAG62293 represent human secreted protein fragments. The genes
XX and their corresponding secreted proteins are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX therapy. Pathological conditions can be diagnosed by determining the
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 52 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of
XX proliferative disorders, cancer, tumours, foetal and developmental
XX abnormalities, haematopoietic disorders, diseases of the immune system,
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX allergies, neurological disorders (e.g., Alzheimer's disease,
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX cardiovascular disorders, angiogenic disorders, kidney disorders,
XX gastrointestinal disorders, pregnancy-related disorders, endocrine
XX disorders, and infections. The proteins can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues, to identify their
XX cognate ligands or binding partners, and in chemotaxis, and can be used
XX as a food additive or preservative to modify storage properties.
XX Antibodies specific for a protein of the invention can be used in
XX alleviating symptoms associated with the disorders mentioned above, and

CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
XX SQ Sequence 2319 BP, 431 A; 696 C; 736 G; 456 T; 0 other;
XX
XX
XX Query Match 56.0%; Score 14; DB 22; Length 2319;
XX Best Local Similarity 100.0%; Pred. No. 65;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 4 tctgctgttgaag 17
XX
XX Db 1993 TCTGCTGTTTGAAG 1980
XX
XX
XX RESULT 39
XX AAD07655/c
XX ID AAD07655 standard; cDNA: 2339 BP.
XX
XX AC AAD07655;
XX
XX 10-AUG-2001 (first entry)
XX
XX
XX Human secreted protein-encoding gene 1 cDNA clone HDPVW11, SEQ ID NO:11.
XX
XX
XX Human: secreted protein; proliferative disorder; cancer; tumour; asthma;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
XX psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
XX inflammation; neurological disorder; Alzheimer's disease; food additive;
XX angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
XX pregnancy-related disorder; endocrine disorder; infection; wound healing;
XX cell culture; chemotaxis; vulnery; binding partner identification;
XX gene therapy; ss.
XX
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 67..1434
XX
XX sig_peptide 67..150 /tag= a
XX /product= "Human secreted protein precursor"
XX
XX mat_peptide 151..1431 /tag= b
XX /product= "Mature human secreted protein"
XX
XX
XX WO200134644-A1.
XX
XX 17-MAY-2001.
XX
XX 08-NOV-2000; 2000MO-US30679.
XX
XX 12-NOV-1999; 99US-0164834.
XX 04-AUG-2000; 2000US-0224007.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Olsen HS, Duan RD, Ebner R:
XX
XX WPI: 2001-329070/34.
XX P-PSDB: AAE03202.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 1; Page 394-395; 499pp; English.
XX
XX AAD07655-AAD07695 represent cDNAs corresponding to 15 human secreted
XX protein genes, and AAE03202-AAE03242 represent the proteins they encode.
XX AAE03243-AAE03280 represent human secreted protein fragments or variants.
XX

CC The invention relates to a novel human fatty acid transporter.
 CC PSEC67 (AAB60388), and to cDNA encoding it (AA27417). PSEC67 is
 CC responsible for the uptake of oleic acid into cells. The invention
 CC also relates to vectors and host cells comprising a PSEC67
 CC nucleic acid; the recombinant production of PSEC67; an antibody
 CC against PSEC67; methods of screening for compounds which can regulate
 CC the uptake of long-chain fatty acids into cells; and the compounds thus
 CC identified. The PSEC67 protein and the gene encoding it are useful as
 CC targets for the treatment or prevention of diseases associated with
 CC defective metabolism of long-chain fatty acids. Such diseases include
 CC cardiomyopathy, skeletal muscle disorders, or renal failure. PSEC67
 CC nucleic acids may also be used in gene therapy for such disorders.
 CC The present sequence represents cDNA encoding PSEC67.

XX Sequence 2405 BP; 410 A; 735 C; 781 G; 479 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2405;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
 |||||
 Db 2111 TCTGCTGTTCGAG 2098

RESULT 42
 AAH19186/c
 ID AAH19186 standard; cDNA; 2447 BP.

XX AAH19186;

DT 17-JUL-2001 (first entry)

DE Human secreted protein-encoding gene 17 cDNA clone HSLCA15, SEQ ID NO:27.

XX Human; secreted protein; proliferative disorder; cancer;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder; tumour;
 KW endocrine disorder; infection; wound healing; vunerary;
 KW cell culture; chemotaxis; food additive;
 KW binding partner identification; ss.

OS Homo sapiens.

PN WO200132910-A2.

PD 10-MAY-2001.

PF 25-OCT-2000; 2000WO-US29362.

PR 29-OCT-1999; 99US-0162240.

PR 30-JUN-2000; 2000US-0215131.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Komatsoulis GA, Birse CE, Ni J, Soppet DR;

DR WPI: 2001-335835/35.

XX N-PSDB: AAG62172.

PT Novel 27 isolated human secreted proteins and polynucleotides encoding
 PT them useful for treating, diagnosing, preventing Alzheimer's disease,
 PT Parkinson's disease, AIDS, rheumatoid arthritis, asthma

PS Claim 4; Page 456-457; 594pp; English.

XX AAH19170-AAH19250 represent cDNAs corresponding to 27 human secreted

CC protein genes, and AAG62156-AAG62235 represent the proteins they encode.
 CC AAG62236-AAG62291 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 52 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., Rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein-encoding cDNA of the invention.

XX Sequence 2447 BP; 444 A; 735 C; 781 G; 487 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2447;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
 |||||
 Db 2111 TCTGCTGTTCGAG 2098

RESULT 43
 AA233977/c
 ID AA233977 standard; cDNA; 2574 BP.

XX AA233977;

DT 07-DEC-1999 (first entry)

DE Human PRO703 nucleotide sequence.

XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
 KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
 KW secreted protein; transmembrane protein; ss.

OS Homo sapiens.

PN WO9946281-A2.

PD 16-SEP-1999.

PF 08-MAR-1999; 99WO-US05028.

PR 10-MAR-1998; 98US-0077450.

PR 11-MAR-1998; 98US-0077632.

PR 11-MAR-1998; 98US-0077641.

PR 11-MAR-1998; 98US-0077649.

PR 12-MAR-1998; 98US-0077791.

PR 13-MAR-1998; 98US-0078004.

PR 17-MAR-1998; 98US-0040220.

PR 20-MAR-1998; 98US-0078886.

PR 20-MAR-1998; 98US-0078910.

PR 20-MAR-1998; 98US-0078936.

[illegible]

```

PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH ) GENENTECH INC.
PI Ashkenazi AJ, Baker KP, Bolstein D, Desnoyers L, Eaton DL,
PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME,
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ,
PI Kijavyn IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA,
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WT;
XX
XX WPI; 2000-611443/58.
DR P-PSDB; AAB44255.
XX
XX Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities .
XX
XX Claim 2; Fig 38; 636pp; English.
XX
XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.
XX
XX Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;
SQ
Query Match 56.0%; Score 14; DB 21; Length 2574;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4 tctgctgttgaag 17
Db 2167 TCTGCTGTTGAAG 2154
RESULT 45
AAC58239/C
ID AAC58239 standard; cDNA; 2574 BP.
XX
XX AAC58239;
XX
XX 25-JAN-2001 (first entry)
XX
XX Human PRO703 nucleotide sequence SEQ ID NO:28.
XX
XX Human; tumour; diagnosis; neoplastic disease; identification; cancer;
XX tumorigenesis; detection; neoplastic cell growth; proliferation;
XX cytostatic; antiinflammatory; immunomodulatory; inflammatory disorder;
XX immunological disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200053754-A1.
XX
XX 14-SEP-2000.
XX
XX 06-JAN-2000; 2000WO-US00277.
XX
XX

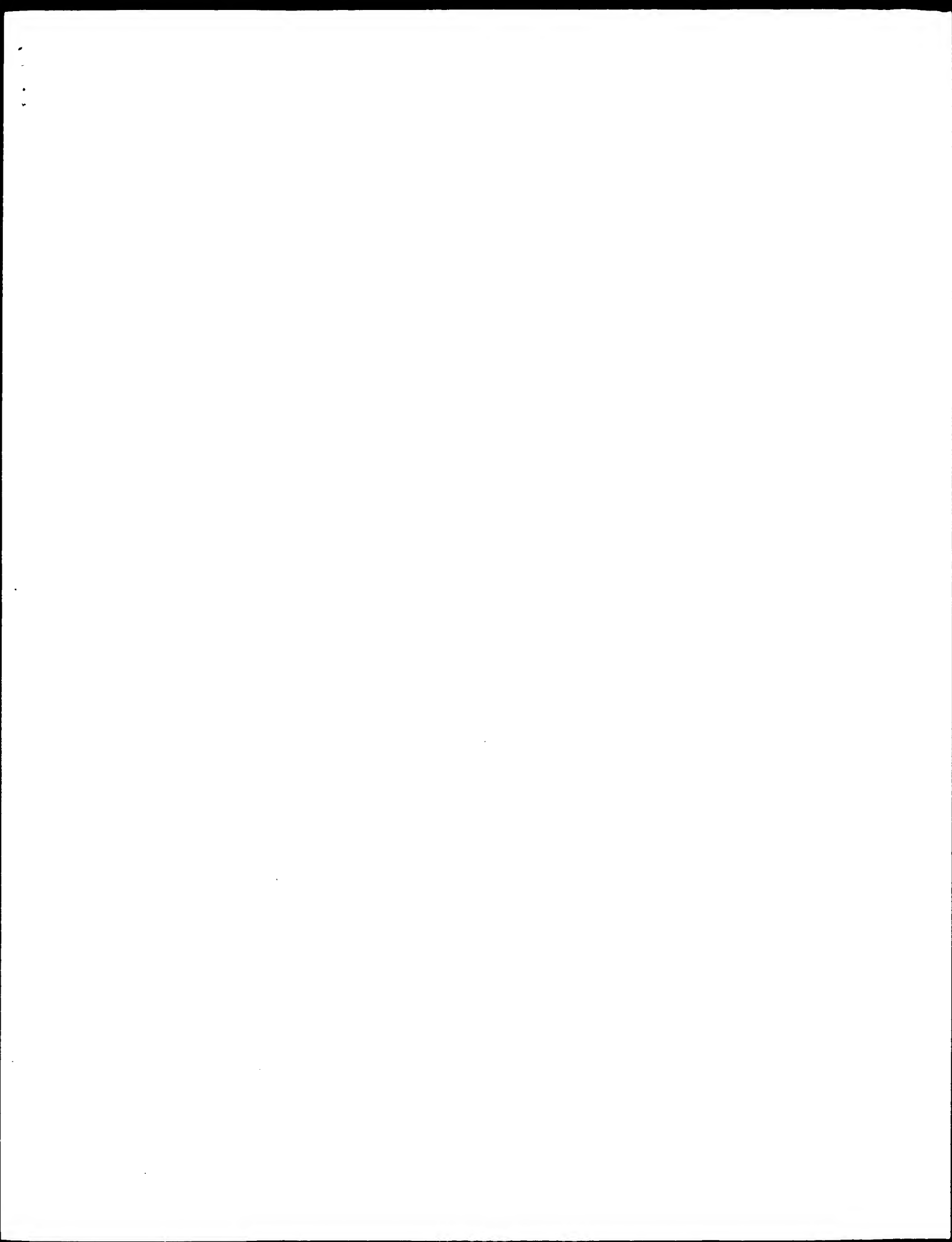
```

```

XX
XX 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 05-OCT-1999; 99WO-US23089.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
XX
XX (GETH ) GENENTECH INC.
PI Baker KP, Desauvage FJ, Goddard A, Gurney AL, Klein RD, Roy MA,
PI Wood WT;
XX
XX WPI; 2000-572269/53.
DR P-PSDB; AAB24054.
XX
XX New isolated antibody for use in compositions and methods for the
PT diagnosis and treatment of neoplastic cell growth and proliferation in
PT mammals, including humans, and in monitoring tumor treatment .
XX
XX Claim 50; Fig 28; 195pp; English.
XX
XX The present invention describes an isolated antibody (Ab) that binds to
CC one of the human proteins (P) designated PRO213, PRO330, PRO1449,
CC PRO337, PRO334, PRO351, PRO362, PRO615, PRO531, PRO538, PRO3664, PRO618,
CC PRO772, PRO703, PRO792 or PRO474. The Ab can be used in compositions
CC and methods for the diagnosis and treatment of neoplastic cell growth
CC and proliferation in mammals, including humans. Genes and polypeptides
CC encoded by them, that are amplified in the genome of a tumour cell, can
CC be identified and are useful targets for the treatment and prevention of
CC certain cancers and may be used to monitor tumour treatment. Compounds
CC that inhibit the expression or activity of the identified polypeptides
CC can be identified and used as antagonists. Benign or malignant tumours,
CC inflammatory disorders and immunological disorders can be treated.
CC AAC58123 to AAC58224 represent hybridisation probes and PCR primers used
CC in the isolation of the human PRO sequences. AAC58225 to AAC58241 and
CC AAB24041 to AAB24056 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX
XX Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;
SQ
Query Match 56.0%; Score 14; DB 21; Length 2574;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4 tctgctgttgaag 17
Db 2167 TCTGCTGTTGAAG 2154

```

Search completed: December 26, 2001, 14:16:40
Job time: 6992 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:52:35 ; Search time 143.1 Seconds
(Without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196f-4
Perfect score: 25
Sequence: 1 ggaatcgtcgttgaagcgcagcag 25

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 11323899 residues

Word size : 9
Total number of hits satisfying chosen parameters: 12481

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*\n2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*\n3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*\n4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*\n5: /cgn2_6/ptodata/2/ina/PCrUS.COMB.seq:*\n6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appl1
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appl1
3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appl1
4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appl1
5	14	56.0	753	4	US-09-232-191-14	Sequence 14, Appl1
6	14	56.0	753	4	US-09-232-200-14	Sequence 14, Appl1
7	14	56.0	753	4	US-09-232-197-14	Sequence 14, Appl1
8	14	56.0	1173	4	US-09-232-200-50	Sequence 50, Appl1
9	14	56.0	1173	4	US-09-232-197-50	Sequence 50, Appl1
10	14	56.0	1998	4	US-09-232-200-68	Sequence 68, Appl1
11	14	56.0	1998	4	US-09-232-197-68	Sequence 68, Appl1
12	14	56.0	2087	4	US-09-232-191-6	Sequence 6, Appl1
13	14	56.0	2087	4	US-09-232-200-6	Sequence 6, Appl1
14	14	56.0	2087	4	US-09-232-197-6	Sequence 6, Appl1
15	14	56.0	2166	4	US-09-232-200-101	Sequence 101, App
16	14	56.0	2166	4	US-09-232-197-101	Sequence 101, App
17	14	56.0	12284	2	US-08-876-991-1	Sequence 1, Appl1
18	14	56.0	12284	2	US-09-059-853-1	Sequence 1, Appl1
19	13	52.0	17	1	US-08-233-608-8	Sequence 8, Appl1
20	13	52.0	17	1	US-08-887-480-8	Sequence 8, Appl1
21	13	52.0	17	2	US-08-722-187-8	Sequence 8, Appl1
22	13	52.0	17	3	PCT-US95-04712-8	Sequence 8, Appl1
23	13	52.0	23	1	US-08-233-608-9	Sequence 9, Appl1
24	13	52.0	23	1	US-08-887-480-9	Sequence 9, Appl1
25	13	52.0	23	2	US-08-722-187-9	Sequence 9, Appl1
26	13	52.0	23	5	PCT-US95-04712-9	Sequence 9, Appl1
27	13	52.0	27	1	US-08-150-331-31	Sequence 31, Appl1

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
: Sequence 7, Application US/08401068
: Patent No. 5859335
: GENERAL INFORMATION:
: APPLICANT: Patton, David
: TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/401,068
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/351,970
: FILING DATE: 08-DEC-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1041 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1038
: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: /product= "biotin synthase"
: OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgcagcag 25
db 63 ggaatcgtctgttgaagcgcagcag 87

RESULT 2
US-08-846-338-7
: Sequence 7, Application US/0846338
: Patent No. 5869719
: GENERAL INFORMATION:
: APPLICANT: Patton, David
: TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/846,338
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1041 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1038
: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: /product= "biotin synthase"
: OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgcagcag 25
db 63 ggaatcgtctgttgaagcgcagcag 87

RESULT 3
US-08-411-768B-1
: Sequence 1, Application US/08411768B
: Patent No. 6083712
: GENERAL INFORMATION:
: APPLICANT: Olwen Birch
: APPLICANT: Johann Brass
: APPLICANT: Martin Fuhrmann
: APPLICANT: Nicholas Shaw
: TITLE OF INVENTION: Biotechnological Method
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Wordperfect
: CURRENT APPLICATION DATA:
```



```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biocA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /aminoacid="aminoacid"
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="biocA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent"
OTHER INFORMATION: /transcriptional_terminator="rho-independent"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
OTHER INFORMATION: /standard_name="rho-independent"
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter plac"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /number=1
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

```

```

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gaaclgclgclgttgaagcgagcag 25
DB 179 GCATCTGCTGTTGAAGCCGACGAG 203

```

```

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pBO30A15-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1154..2308
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start=1154
; OTHER INFORMATION: /product="KAPA synthase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /gene="bioF"
; OTHER INFORMATION: /number=2
; OTHER INFORMATION: /standard_name="8-Amino-7-oxononanoate synthase"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3043..3753
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start=3043
; OTHER INFORMATION: /EC_number=6.3.3.3
; OTHER INFORMATION: /product="DPA synthase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /gene="bioD"
; OTHER INFORMATION: /number=4
; OTHER INFORMATION: /standard_name="Dethiobiotin synthase"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 1141..1156
; OTHER INFORMATION: /standard_name="bioF RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3030..3045
; OTHER INFORMATION: /standard_name="bioD RBS"
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
; US-08-411-768B-6

Query Match          100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgtttgaagcgaagcag 25
|||||
Db 179 GCATCTGCTGTTGAAGCGACGACG 203

RESULT 5
US-09-232-191-14/c
; Sequence 14, Application US/09232191
; Patent No. 6284487
```

```

; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: Fatty Acid Transport Proteins
; FILE REFERENCE: WH197-21P3ME
; CURRENT APPLICATION NUMBER: US/09/232,191
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-232-191-14
```

```

Query Match          56.0%; Score 14; DB 4; Length 753;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcctgctgttgaag 17
|||||
Db 436 TCTGCTGTTTGAAG 423
```

```

RESULT 6
US-09-232-200-14/c
; Sequence 14, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tarragila, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21P3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-232-200-14
```

```

Query Match          56.0%; Score 14; DB 4; Length 753;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcctgctgttgaag 17
|||||
Db 436 TCTGCTGTTTGAAG 423
```

```

RESULT 7
US-09-232-197-14/c
; Sequence 14, Application US/09232197A
```

```
Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WH197-21P3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
CURRENT FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 753
TYPE: DNA
ORGANISM: Homo sapiens
US-09-232-197-14
```

```
Query Match          56.0%; Score 14; DB 4; Length 753;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 tctgctgttgaag 17
      |||
DB 436 TCTGCTGTTGAAG 423
```

```
RESULT 8
US-09-232-200-50/c
Sequence 50, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WH197-21P3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 50
LENGTH: 1173
TYPE: DNA
ORGANISM: Homo sapiens
US-09-232-200-50
```

```
Query Match          56.0%; Score 14; DB 4; Length 1173;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 tctgctgttgaag 17
      |||
DB 853 TCTGCTGTTGAAG 840
```

```
RESULT 9
US-09-232-197-50/c
Sequence 50, Application US/09232197A
Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WH197-21P3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 50
LENGTH: 1173
TYPE: DNA
ORGANISM: Homo sapiens
US-09-232-197-50
```

```
Query Match          56.0%; Score 14; DB 4; Length 1173;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 tctgctgttgaag 17
      |||
DB 853 TCTGCTGTTGAAG 840
```

```
RESULT 10
US-09-232-200-68/c
Sequence 68, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WH197-21P3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 68
LENGTH: 1998
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-200-68
```

```
Query Match          56.0%; Score 14; DB 4; Length 1998;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 tctgctgttgaag 17
      |||
```

Db 1690 TCTGCTGTTGAAG 1677

```
RESULT 11
US-09-232-197-68/c
; Sequence 68, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21P3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 1998
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-197-68
```

Query Match 56.0%; Score 14; DB 4; Length 1998;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
|||||

Db 1690 TCTGCTGTTGAAG 1677

```
RESULT 12
US-09-232-191-6/c
; Sequence 6, Application US/09232191
; Patent No. 6284487
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: Fatty Acid Transport Proteins
; FILE REFERENCE: WH197-21P3ME
; CURRENT APPLICATION NUMBER: US/09/232,191
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2087
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-191-6
```

Query Match 56.0%; Score 14; DB 4; Length 2087;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 tctgctgttgaag 17

Db 1779 TCTGCTGTTGAAG 1766
|||||

```
RESULT 13
US-09-232-200-6/c
; Sequence 6, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21P3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2087
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-200-6
```

Query Match 56.0%; Score 14; DB 4; Length 2087;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
|||||

Db 1779 TCTGCTGTTGAAG 1766

```
RESULT 14
US-09-232-197-6/c
; Sequence 6, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21P3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2087
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-197-6
```

Query Match 56.0%; Score 14; DB 4; Length 2087;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 14: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
OY 4 tctgcgtttgaag 17
|||||
Db 1779 TCTGCTGTTTGAG 1766

RESULT 15
US-09-232-200-101/C
Sequence 101, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 101
LENGTH: 2166
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (19)...(2124)
US-09-232-200-101

Query Match 56.0%; Score 14; DB 4: Length 2166;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0: Mismatches 0: Indels 0: Gaps 0:
OY 4 tctgcgtttgaag 17
|||||
Db 1987 TCTGCTGTTTGAG 1974

RESULT 16
US-09-232-197-101/C
Sequence 101, Application US/09232197A
Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 101
LENGTH: 2166
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (19)...(2124)
US-09-232-197-101

Query Match 56.0%; Score 14; DB 4: Length 2166;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 4 tctgcgtttgaag 17
|||||
Db 1987 TCTGCTGTTTGAG 1974

RESULT 17
US-08-876-991-1
Sequence 1, Application US/08876991
Patent No. 5925360
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J Igen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESS: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,991
FILING DATE: 16-JUN-1997
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/747,577
FILING DATE:
APPLICATION NUMBER: US/08/650,584
FILING DATE:
APPLICATION NUMBER: US/08/469,702
FILING DATE:
APPLICATION NUMBER: US/08/123,596
FILING DATE:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D

FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_KDA_protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
US-08-876-991-1

Query Match 56.0%; Score 14; DB 2; Length 12284;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
|||||
Db 2253 TCTGCTGTTGAAG 2266

RESULT 18
US-09-059-853-1
Sequence 1, Application US/09059853
Patent No. 5935582
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Helinz-J igen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEO ID NO: 1;
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Hog cholera virus
STRAIN: Allort
CELL LINE: PK 15 and 38A1D
FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_KDA_protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
US-09-059-853-1

Query Match 56.0%; Score 14; DB 2; Length 12284;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
|||||
Db 2253 TCTGCTGTTGAAG 2266

RESULT 19
US-08-233-608-8/c
Sequence 8, Application US/08233608
Patent No. 5585238
GENERAL INFORMATION:
APPLICANT: Ligon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,608
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spurrill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615

TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid
DESCRIPTION: Oligonucleotide primer JB434
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-233-608-8

Query Match 52.0%; Score 13; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcagca 24
|||||
DB 15 TTGAAGCGACGA 3

RESULT 20
US-08-887-480-8/C
Sequence 8, Application US/08887480
PATENT No. 5814453
GENERAL INFORMATION:
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
TITLE OF INVENTION: Polymerase Chain Reaction
NUMBER OF SEQUENCES: 96
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5814453artis Corporation
STREET: 520 White Plains Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,480
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,187
FILING DATE: 15-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1739/PCT/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid
DESCRIPTION: Oligonucleotide primer JB434
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-887-480-8

Query Match 52.0%; Score 13; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcagca 24
|||||
DB 15 TTGAAGCGACGA 3

RESULT 21
US-08-722-187-8/C
Sequence 8, Application US/08722187
PATENT No. 5953274
GENERAL INFORMATION:
APPLICANT: Ligon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
TITLE OF INVENTION: Polymerase Chain Reaction
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,187
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,608
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Walsh, Andrea C.
REGISTRATION NUMBER: 34,988
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8666
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid
DESCRIPTION: Oligonucleotide primer JB434
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-722-187-8

Query Match 52.0%; Score 13; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcagca 24
|||||
DB 15 TTGAAGCGACGA 3

RESULT 22
PCT-US95-04712-8/C
Sequence 8, Application PC/TUS9504712
GENERAL INFORMATION:
APPLICANT: Ligon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
TITLE OF INVENTION: Polymerase Chain Reaction

NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04712
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,608
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Walsh, Andrea C.
REGISTRATION NUMBER: 34,988
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8666
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: Oligonucleotide primer JB434
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-04712-8

Query Match 52.0%; Score 13; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgacga 24
|||||
Db 15 TTGAAGCGCAGCA 3

RESULT 23
US-08-233-608-9/c
Sequence 9, Application US/08233608
Patent No. 5585238
GENERAL INFORMATION:
APPLICANT: Ligon, James M
TITLE OF INVENTION: Detection of Fungal Pathogens using the
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,608

FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid
DESCRIPTION: Oligonucleotide primer JB525
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-233-608-9

Query Match 52.0%; Score 13; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 24
US-08-887-480-9/c
Sequence 9, Application US/08887480
Patent No. 5814453
GENERAL INFORMATION:
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens using the
NUMBER OF SEQUENCES: 96
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5814453artis Corporation
STREET: 520 White Plains Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,480
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,187
FILING DATE: 15-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1739/PCT/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid

DESCRIPTION: Oligonucleotide primer JB525
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-887-480-9

Query Match 52.0%; Score 13; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 25
US-08-722-187-9/c
; Sequence 9, Application US/08722187
; Patent No. 5955274
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,187
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,608
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Walsh, Andrea C.
; REGISTRATION NUMBER: 34,988
; REFERENCE/DOCKET NUMBER: CGC 1739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8666
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Oligonucleotide primer JB525
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-722-187-9

Query Match 52.0%; Score 13; DB 2; Length 23;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 26
PCT-US95-04712-9/c
; Sequence 9, Application PC/TUS9504712
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04712
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,608
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Walsh, Andrea C.
; REGISTRATION NUMBER: 34,988
; REFERENCE/DOCKET NUMBER: CGC 1739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8666
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Oligonucleotide primer JB525
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-04712-9

Query Match 52.0%; Score 13; DB 5; Length 23;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 27
US-08-150-331-31/c
; Sequence 31, Application US/08150331
; Patent No. 5516512
; GENERAL INFORMATION:
; APPLICANT: DORSSERS J., LAMBERTUS C.
; APPLICANT: VAN LEEN, ROBERT W.
; TITLE OF INVENTION: MUTANTS OF HUMAN INTERLEUKIN-3
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/150,331
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/651,437
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: GRACEY, NANCY J.
REGISTRATION NUMBER: 28,216
REFERENCE/DOCKET NUMBER: 24615-20010.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-150-331-31

```

```

Query Match          52.0%; Score 13; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 12 ttgaagcgacga 24
    |||||||
Db 20 TTGAAGCGACGA 8

```

```

RESULT 28
US-08-820-170A-8/c
Sequence 8, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid

```

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
US-08-820-170A-8

```

```

Query Match          52.0%; Score 13; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggaatcgcgtgtt 13
    |||||||
Db 349 GGATCTGCTGTT 337

```

```

RESULT 29
US-09-055-699-8/c
Sequence 8, Application US/09055699
Patent No. 6005088
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,170
FILING DATE:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
US-09-055-699-8

```

```

Query Match          52.0%; Score 13; DB 3; Length 384;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggaatcgcgtgtt 13
    |||||||
Db 349 GGATCTGCTGTT 337

```

```

RESULT 30
US-09-273-565-8/c
Sequence 8, Application US/09273565A
Patent No. 6166190
GENERAL INFORMATION:

```

```

? APPLICANT: FUJIMURA, TSUTOMU
? APPLICANT: MATANABE, TAKESHI
? APPLICANT: HORIE, MASATO
? TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
? TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
? FILE REFERENCE: 0-535599
? CURRENT APPLICATION NUMBER: US/09/273,565A
? EARLIER FILING DATE: 1999-03-22
? EARLIER APPLICATION NUMBER: 09/055,699
? EARLIER FILING DATE: 1998-04-07
? EARLIER APPLICATION NUMBER: 08/820,170
? EARLIER FILING DATE: 1997-03-19
? EARLIER APPLICATION NUMBER: JP 63410/1996
? EARLIER FILING DATE: 1996-03-19
? EARLIER APPLICATION NUMBER: JP 69163/1997
? EARLIER FILING DATE: 1997-03-05
? NUMBER OF SEQ ID NOS: 95
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO: 8
? LENGTH: 384
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-273-565-8

Query Match      52.0%; Score 13; DB 4; Length 384;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  ggaatcgtctgtt 13
Db      349  ggaatcgtctgtt 337

RESULT 31
US-08-233-608-2
; Sequence 2, Application US/08233608
; Patent No. 5585238
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,608
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprull, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: CGC 1739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8615
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

? MOLECULE TYPE: DNA (genomic)
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Septoria nodorum
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1..583
? OTHER INFORMATION: /note= "DNA sequence for the
? US-233-608-2 Internal Transcribed Spacer of Septoria nodorum"

Query Match      52.0%; Score 13; DB 1; Length 583;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      12  ttgaagcgagca 24
Db      464  ttgaagcgagca 476

RESULT 32
US-08-887-480-2
; Sequence 2, Application US/08887480
; Patent No. 5814453
; GENERAL INFORMATION:
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5814453artis Corporation
; STREET: 520 White Plains Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,480
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,187
; FILING DATE: 15-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1739/PCT/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Septoria nodorum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..30
; OTHER INFORMATION: /note= "3' end of small subunit
; OTHER INFORMATION: rRNA gene"

```

```
FEATURE:
NAME/KEY: misc.feature
LOCATION: 31..216
OTHER INFORMATION: /note= "ITS 1"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 217..372
OTHER INFORMATION: /note= "5.8S rRNA gene"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 373..526
OTHER INFORMATION: /note= "ITS 2"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 527..583
OTHER INFORMATION: /note= "5' end of large subunit"
OTHER INFORMATION: rRNA gene"
US-08-887-480-2
```

```
Query Match          52.0%; Score 13; DB 1; Length 583;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 12 ttgaagcgacga 24
      |||
Db 464 TTGAAGCGCAGCA 476
```

```
RESULT 33
US-08-722-187-2
Sequence 2, Application US/08722187
Patent No. 5955274
GENERAL INFORMATION:
APPLICANT: Ligon, James M
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
NUMBER OF INVENTION: Polymerase Chain Reaction
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,187
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,608
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Walsh, Andrea C.
REGISTRATION NUMBER: 34,988
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8666
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 583 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
```

```
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Septoria nodorum
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..583
OTHER INFORMATION: /note= "DNA sequence for the
US-08-722-187-2 Internal Transcribed Spacer of Septoria nodorum"
```

```
Query Match          52.0%; Score 13; DB 2; Length 583;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 12 ttgaagcgacga 24
      |||
Db 464 TTGAAGCGCAGCA 476
```

```
RESULT 34
PCT-US95-04712-2
Sequence 2, Application PC/TUS9504712
GENERAL INFORMATION:
APPLICANT: Ligon, James M
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
NUMBER OF INVENTION: Polymerase Chain Reaction
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04712
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,608
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Walsh, Andrea C.
REGISTRATION NUMBER: 34,988
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8666
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 583 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Septoria nodorum
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..583
OTHER INFORMATION: /note= "DNA sequence for the
PCT-US95-04712-2 Internal Transcribed Spacer of Septoria nodorum"
```

Query Match 52.0%; Score 13; DB 5; Length 583;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcagca 24
|||||
DB 464 TTGAAGCAGCAGCA 476

RESULT 35

US-08-998-416-727
Sequence 727, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYP11
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6239264arlis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NO. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 727:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1475RP
US-08-998-416-727

Query Match 52.0%; Score 13; DB 4; Length 749;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcctttgaagc 18
|||||
DB 96 TGCTGTTTGAAGC 108

RESULT 36

US-08-153-848-31
Sequence 31, Application US/08153848
Patent No. 5759804
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESS: Bicknell
STREET: 6300 Sears Tower, 213 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,848
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5759804and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: exon
LOCATION: 7..80
FEATURE:
NAME/KEY: CDS
LOCATION: 94..1158
US-08-153-848-31

Query Match 52.0%; Score 13; DB 1; Length 1161;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatcgcctgctt 14
|||||
DB 322 GATCTGCTGCTTGG 334

RESULT 37

PCT-0593-11153-31
Sequence 31, Application PC/TUS9311153
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: Novel Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

ADDRESSER: Bicknell
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11153
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Noland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: exon
LOCATION: 7..80
FEATURE:
NAME/KEY: CDS
LOCATION: 94..1158
PCT-US93-11153-31

Query Match 52.0%; Score 13; DB 5; Length 1161;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgttg 14
|||||

DB 322 GATCTGCTGTTG 334

RESULT 38
US-08-252-966B-13/c
Sequence 13, Application US/08252966B
Patent No. 5624818
GENERAL INFORMATION:
APPLICANT: Eisenman, Robert N.
APPLICANT: Hurlin, Peter J.
APPLICANT: Ayer, Donald E.
TITLE OF INVENTION: Regulatory Proteins that Dimerize with
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson, and Kindness PLLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/252,966B
FILING DATE: 01-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: FHCRI7694
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-8100
TELEFAX: (206) 224-0779
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1242 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: clone 10; see Figure 24
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
US-08-252-966B-13

Query Match 52.0%; Score 13; DB 1; Length 1242;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgcctgttgag 17
|||||

DB 525 CTGCTGTTGAG 513

RESULT 39
US-08-353-550-7/c
Sequence 7, Application US/08353550
Patent No. 5744313
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Cavanaugh, William M.
TITLE OF INVENTION: No. 5744313el Protein Domain which Binds
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Hourie and Crew
STREET: One Market Plaza, Stewart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,550
FILING DATE: 09-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 02307K-057300
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1422
US-08-353-550-7

Query Match 52.0%; Score 13; DB 1; Length 1425;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 tgttgaagcgca 21
|||||
Db 611 TGTGGAAGCGCA 599

RESULT 40
US-08-551-687-7/C
Sequence 7, Application US/08551687
Patent No. 5925547
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Cavanaugh, William M.
TITLE OF INVENTION: No. 5925547el Protein Domain Which Binds
TITLE OF INVENTION: Tyrosine Phosphorylated Proteins
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew
STREET: One Market Plaza, Stewart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/551,687
FILING DATE: 01-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/353,550
FILING DATE: 09-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Matthew B.
REGISTRATION NUMBER: P39,787
REFERENCE/DOCKET NUMBER: 2307K-5731
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1422
US-08-551-687-7

Query Match 52.0%; Score 13; DB 2; Length 1425;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 tgttgaagcgca 21
|||||
Db 611 TGTGGAAGCGCA 599

RESULT 41
US-08-820-170A-9/C
Sequence 9, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIMURA
APPLICANT: Takeshi, MATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
LIBRARY: Human fetal brain CDNA library
CLONE: GEN-025F07
FEATURE:
NAME/KEY: CDS
LOCATION: 95..478
US-08-820-170A-9

Query Match 52.0%; Score 13; DB 2; Length 1493;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgatgtt 13
|||||
Db 443 GCATCTGCTGTTT 431

RESULT 42
US-09-055-699-9/C
Sequence 9, Application US/09055699
Patent No. 6005088
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIMURA
APPLICANT: Takeshi, MATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas

```

: STREET: 2100 Pennsylvania Avenue, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: United States
: ZIP: 20037-3202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/055,699
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: 08/820,170
: FILING DATE:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 293-7060
: TELEFAX: (202) 293-7860
: TELEX: 6491103
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1493 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA(genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: IMMEDIATE SOURCE:
: LIBRARY: Human fetal brain cDNA library
: CLONE: GEN-025F07
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 95..478
: US-09-055-699-9

Query Match          52.0%; Score 13; DB 3; Length 1493;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatgcgtctt 13
    |||
Db 443 GGATCTGCTGTTT 431

RESULT 43
: US-09-273-565-9/c
: Sequence 9, Application US/09273565A
: Patent No. 6166190
: GENERAL INFORMATION:
: APPLICANT: FUJIMURA, TSUTOMU
: APPLICANT: MATANABE, TAKESHI
: APPLICANT: HORIE, MASATO
: TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
: FILE REFERENCE: O-53599
: CURRENT APPLICATION NUMBER: US/09/273,565A
: CURRENT FILING DATE: 1999-03-22
: EARLIER APPLICATION NUMBER: 09/055,699
: EARLIER FILING DATE: 1998-04-07
: EARLIER APPLICATION NUMBER: 08/820,170
: EARLIER FILING DATE: 1997-03-19
: EARLIER APPLICATION NUMBER: JP 63410/1996
: EARLIER FILING DATE: 1996-03-19
: EARLIER APPLICATION NUMBER: JP 69163/1997
: EARLIER FILING DATE: 1997-03-05
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 9
: LENGTH: 1493

```

```

: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (95)..(478)
: US-09-273-565-9

Query Match          52.0%; Score 13; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatgcgtctt 13
    |||
Db 443 GGATCTGCTGTTT 431

RESULT 44
: US-08-171-299B-1
: Sequence 1, Application US/08171299B
: Patent No. 5599665
: GENERAL INFORMATION:
: APPLICANT: Barbieri, Joseph T.
: APPLICANT: Frank, Dara W.
: TITLE OF INVENTION: EXONZYME S PROTEIN PREPARATION AND CLONED
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: QUARLES & BRADY
: STREET: 411 East Wisconsin Avenue
: CITY: Milwaukee
: STATE: Wisconsin
: COUNTRY: U.S.A.
: ZIP: 53202-4497
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/171,299B
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Baker, Jean C.
: REGISTRATION NUMBER: 35,433
: REFERENCE/DOCKET NUMBER: 650053,90871
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (414) 277-5709
: TELEFAX: (414) 271-3552
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1712 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-171-299B-1

Query Match          52.0%; Score 13; DB 1; Length 1712;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgcagcag 25
    |||
Db 388 TGAAGCGCAGCAG 400

RESULT 45
: US-09-221-235-9/c
: Sequence 9, Application US/09221235
: Patent No. 6043040
: GENERAL INFORMATION:

```



```

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,235
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1743
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1743)
US-09-221-235-9

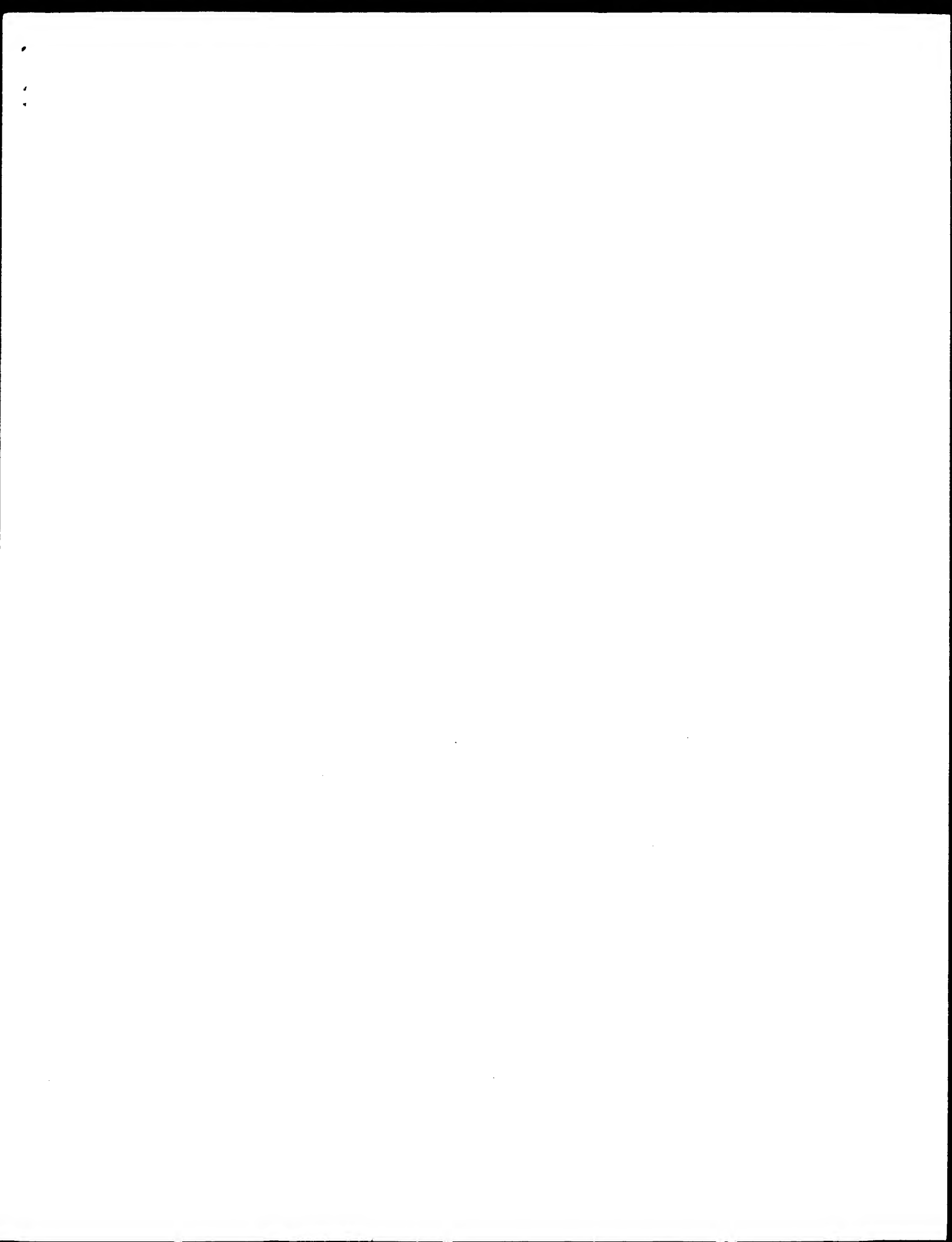
```

```

Query Match      52.0%; Score 13; DB 3; Length 1743;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 13 tgaagcgcagcag 25
   ||| ||| ||| |||
Db 64 TGAAGCCGCGACAG 52

```

Search completed: December 26, 2001, 12:52:37
 Job time: 6789 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:10:04 : Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196F-4

Perfect score: 25

Sequence: 1 ggaatcgtcttgaagcgcagcag 25

Scoring table: OLIGO_NUC
Gapex 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 9

Total number of hits satisfying chosen parameters: 752014

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

EST:*
1: em_esthm:*
2: em_esthm:*
3: em_esthm:*
4: em_esthm:*
5: em_esthm:*
6: em_esthm:*
7: em_esthm:*
8: em_esthm:*
9: em_esthm:*
10: em_esthm:*
11: em_esthm:*
12: em_esthm:*
13: em_esthm:*
14: em_esthm:*
15: em_esthm:*
16: em_esthm:*
17: em_esthm:*
18: em_esthm:*
19: em_esthm:*
20: em_esthm:*
21: em_esthm:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	68.0	381	11	BG888762 EST514613
2	17	68.0	479	11	BE508108 ds26109.y
3	16	64.0	293	13	AZ214122 Sheared D
4	16	64.0	358	11	AZ214122 Sheared D
5	16	64.0	360	11	C61070
6	16	64.0	502	10	BE515828 WHE0606_B
7	16	64.0	517	13	AO692054 HS_5408_A
8	16	64.0	555	13	AO889100 HS_2204_A
9	16	64.0	558	10	AM064984
10	16	64.0	567	10	BE232412
11	16	64.0	632	13	AZ449344
12	16	64.0	650	13	TA69C05P

13	15	60.0	177	13	FR0044714
14	15	60.0	190	11	BI024308
15	15	60.0	278	10	AI021028
16	15	60.0	298	10	BB337949
17	15	60.0	316	13	FR0039786
18	15	60.0	358	11	BF355075
19	15	60.0	372	13	AO521765
20	15	60.0	406	13	AO143500
21	15	60.0	424	13	AO494880
22	15	60.0	427	13	BM077581
23	15	60.0	435	13	AO114851
24	15	60.0	437	10	AM432846
25	15	60.0	456	11	BF555582
26	15	60.0	466	10	AI477772
27	15	60.0	479	13	AZ965014
28	15	60.0	495	13	AO192174
29	15	60.0	507	10	BE114310
30	15	60.0	509	10	AM915326
31	15	60.0	528	10	AA818607
32	15	60.0	532	13	TA375E1LP
33	15	60.0	542	10	AA790463
34	15	60.0	559	13	TA162C10P
35	15	60.0	594	13	FR0036852
36	15	60.0	598	13	FR0035666
37	15	60.0	600	13	FR0044579
38	15	60.0	611	13	FR0044632
39	15	60.0	621	11	BC306774
40	15	60.0	627	11	BC017095
41	15	60.0	632	13	AO754510
42	15	60.0	640	11	BS456580
43	15	60.0	644	11	BS6969048
44	15	60.0	645	11	BE597936
45	15	60.0	649	11	BC085090
46	15	60.0	663	11	AM633493
47	15	60.0	681	11	BE820949
48	15	60.0	716	11	BE971333
49	15	60.0	748	10	AM128741
50	15	60.0	752	13	CNS02060
51	15	60.0	846	11	BC428569
52	15	60.0	870	13	CNS04AGY
53	15	60.0	878	11	BC721419
54	15	60.0	888	13	CNS04KH2
55	15	60.0	893	13	CNS04YX0
56	15	60.0	922	13	CNS02F00
57	15	60.0	961	13	CNS01150
58	15	60.0	966	13	CNS01110
59	15	60.0	977	13	CNS05364
60	15	60.0	1044	13	CNS05B81
61	15	60.0	138	11	BE189673
62	15	60.0	139	11	AZ613372
63	15	60.0	144	11	BE955843
64	15	60.0	161	11	BI389308
65	15	60.0	205	10	AA884873
66	15	60.0	211	10	BE456581
67	15	60.0	216	10	BE200491
68	15	60.0	227	10	BE012248
69	15	60.0	229	11	BE369300
70	15	60.0	230	13	AO506129
71	15	60.0	236	11	D24951
72	15	60.0	241	10	AV040895
73	15	60.0	247	11	BE217212
74	15	60.0	251	10	BE601253
75	15	60.0	261	10	AV360182
76	15	60.0	268	10	AV002603
77	15	60.0	271	11	BC308817
78	15	60.0	281	10	BI882699
79	15	60.0	282	11	BE182292
80	15	60.0	285	13	AZ084513
81	15	60.0	288	10	BI125447
82	15	60.0	300	11	NA7165
83	15	60.0	303	10	BE276182
84	15	60.0	305	10	AV136350
85	15	60.0	307	11	BC439729

AL132206	Fugu	rubr
BI024308	CMS-MT029	
AI021028	ua99001.r	
BB337949	BB337949	
AL127281	Fugu	rubr
BF355075	KC2-HT082	
AO521765	HS-5239_B	
AO143500	HS-3066_B	
AO494880	HS-5195_A	
BM077581	RPC1-24-2	
AO114851	CIT-HSP-2	
AM432846	SH1905.Y	
BF555582	UT-R-E0-C	
AI477772	FD54h11.Y	
AZ965014	2M0234K24	
AO192174	HS-3227_B	
BE114310	UT-R-B1-L	
AM915326	EST346630	
AA818607	UT-R-A0-B	
AL495591	T. brucei	
AA790463	vw04q12.r	
AL472440	T. brucei	
AL124355	Fugu	rubr
AL132071	Fugu	rubr
AL132183	Fugu	rubr
AL132174	Fugu	rubr
BC306774	fm06h08.Y	
BC017095	dfe8903.x	
AO754510	HS-5434_A	
BS456580	602574435	
BS6969048	602835003	
BE597936	P11_67-C0	
BC085090	H3108E05-	
AM633493	bl08d10.W	
BE820949	GM700013A	
BE971333	602839869	
AM128741	fe36h03.Y	
AL206315	Tetradon	
BC428569	602501093	
AL289627	Tetradon	
BC721419	602694974	
AL294815	Tetradon	
AL319405	Tetradon	
AL319405	Tetradon	
AL195369	Tetradon	
AL145285	Anopheles	
AL165959	Tetradon	
AL319405	Tetradon	
AL329499	Tetradon	
BI8189673	RSTR719_A	
AZ613372	IM0441M11	
BE955843	PM2-NN120	
BI389308	949052E12	
AA884873	am21c11.s	
BE456581	ut11c03.x	
BE200491	BE200491	
BE012248	945016A03	
AO506129	CM4-GN008	
AO506129	RPC1-11-3	
D24951	RICR2836A_R	
AV040895	AV040895	
BE217212	RST36914	
BB601253	BB601253	
AV360182	AV360182	
AV002603	AV002603	
BC308817	f189d01.x	
BI882699	EST1157_A	
BE182292	RPC1-23-3	
AZ084513	RPC1-23-3	
BI125447	BI125447	
NA7165	yy73c07.r1	
BE276182	BB276182	
AV136350	AV136350	
BC439729	dab26d02..	

```

86 14 56.0 312 10 AA623286 v119f05.f
87 14 56.0 313 11 B1188823
88 14 56.0 316 10 BB154393
89 14 56.0 322 10 BB384422
90 14 56.0 332 11 BG198869
91 14 56.0 334 11 BG193289
92 14 56.0 337 10 AM127031
93 14 56.0 343 10 AM245381
94 14 56.0 345 10 A1708068
95 14 56.0 345 10 BB165235
96 14 56.0 347 10 A1302370
97 14 56.0 348 10 AA902490
98 14 56.0 362 10 A1669961
99 14 56.0 362 10 A1810191
c 100 14 56.0 363 10 AA418282

```

ALIGNMENTS

```

RESULT 1
LOCUS BC888762/c
DEFINITION BC888762 381 bp mRNA EST 30-MAV-2001
ACCESSION E57514613 cSTD Solanum tuberosum cDNA clone cSTD11014.5' sequence,
VERSION BC888762
KEYWORDS BC888762.1 GI:14265848
SOURCE EST.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 381)
van der Hoeven, R., Bezzerides, J., Ewing, E., Cho, J., Chiemling, A.,
Bongri, O., Buell, C. R., Ronning, C., Tanksley, S. and Baker, B.
Generations of ESTs from dormant potato tubers
Unpublished (2001)
Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cnaa@resgen.com
Seq primer: M13P-R.

```

FEATURES

```

SOURCE
Location/Qualifiers
1..381
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTD11014"
/clone_lib="cSTD"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/note="Vector: Bluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; This library targets genes expressed in dormant
tubers. This library was made from sections of dormant
tuber, avoiding the buds and epidermis. Tubers were stored
for one month post-harvest at 40C. The tuber was peeled,
well away from the surface. Then it was chopped into 1-2
mm cubes and immediately frozen in liquid nitrogen. This
library is noted as p4 in Tanksley lab notebooks."
BASE COUNT 88 a 76 c 111 g 106 t
ORIGIN

```

```

Query Match 68.0%; Score 17; DB 11; Length 381;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 tgttgaagcgagcgag 25
|||||
DB 311 TGTTTGAGCGAGCAG 295

```

```

RESULT 2
LOCUS BE508108/c
DEFINITION BE508108 479 bp mRNA EST 07-AUG-2000
dc26f09.y1 NICHD XGC l11 Xenopus laevis cDNA clone IMAGE:3398249.5'
similar to TR:075940 075940 30KDA SPLICING FACTOR. ; mRNA
sequence.
BE508108.1 GI:9727883
VERSION BE508108
KEYWORDS African clawed frog.
SOURCE EST.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 479)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LNL at: info@image.lnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 434.

```

FEATURES

```

SOURCE
Location/Qualifiers
1..479
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3398249"
/clone_lib="NICHD XGC l11"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Liver; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 Kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC)
Library."
BASE COUNT 114 a 147 c 112 g 106 t
ORIGIN

```

```

Query Match 68.0%; Score 17; DB 10; Length 479;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 tgttgaagcgagcgag 25
|||||
DB 183 TGTTTGAGCGAGCAG 167

```

```

RESULT 3
LOCUS A2214122
DEFINITION A2214122 293 bp DNA GSS 09-JUN-2000
Sheared DNA-67G1.TF Sheared DNA Trypanosoma brucei genomic clone
A2214122
ACCESSION A2214122
VERSION A2214122.1 GI:8431922
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 293)
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Fraser, C. and Adams, M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library

```

JOURNAL COMMENT

Unpublished (1999)
 Other:GSSS: Sheared DNA-67G1.TR
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 0208
 Email: nelsayed@tigr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through Research Genetics, Alabama, USA. Sheared DNA end sequences search page: <http://www.tigr.org/cdb/mdb/tbdb/>.
 Seq primer: M13-Forward
 Class: Shotgun.

FEATURES

Location/Qualifiers
 1..293
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone_lib="Sheared DNA-67G1"
 /note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 38 a 73 c 78 g 104 t
 ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 293;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctcgtcttgaagc 18
 |||||||
 Db 149 ATCTGCTGTTGAAGC 164

RESULT 4
 LOCUS AI207085 358 bp mRNA EST 10-NOV-1998
 DEFINITION q159609.x1 Soares, testis_NHT Homo sapiens cDNA clone IMAGE:1754296
 3' similar to TR:061655 061655 EUKARYOTIC TRANSLATION INITIATION
 FACTOR 4A RELATED SEQUENCE 1; mRNA sequence.

ACCESSION AI207085 GI:3765757
 VERSION AI207085.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 358)
 AUTHORS Mammali, Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsofremail.nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 , Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILN at:
www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 652 Std Error: 0.00
 Seq primer: -400p from Gibco
 High quality sequence stop: 239.
 Location/Qualifiers

FEATURES

1..358
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1754296"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc. and primed with a Not I - oligo(dT) primer [5', TGTTACCATCTGGAAGTGGAGCGCCGCCCAATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 98 a 79 c 73 g 107 t
 ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 358;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ctgttgaagcgcagc 23
 |||||||
 Db 294 CTGTTGAAGCGCAGC 309

RESULT 5
 LOCUS C61070 360 bp mRNA EST 22-SEP-1997
 DEFINITION C61070 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
 clone yK224c7 5', mRNA sequence.

ACCESSION C61070 GI:2419775
 VERSION C61070.1
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.

REFERENCE 1 (bases 1 to 360)
 AUTHORS Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano
 , M., Miyata, A. and Nishigaki, A.
 TITLE Expression map of the C.elegans genome
 JOURNAL Unpublished (1996)
 COMMENT Contact: Yuji Kohara
 Genome Biology Lab.
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.

FEATURES
 source
 1..360
 /organism="Caenorhabditis elegans"
 /strain="CH1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="yK224c7"
 /clone_lib="Yuji Kohara unpublished cDNA"
 /sex="hermaphrodite, male"
 /tissue_type="whole animal"
 /dev_stage="varied"

BASE COUNT 101 a 71 c 76 g 111 t 1 others
 ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 360;

Best Local Similarity 100.0%; Pred. No. 68;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgtctgttgaagc 18
|||||
Db 26 ATCTGCTGTTGAAGC 11

RESULT 6

LOCUS BE515828 502 bp mRNA EST 08-AUG-2000
DEFINITION WHE0606_B07_C14ZA Wheat ABA-treated embryo cDNA library Triticum
aestivum cDNA clone WHE0606_B07_C14, mRNA sequence.

ACCESSION BE515828
VERSION BE515828.1 GI:9739842
KEYWORDS EST.

SOURCE Triticum aestivum
ORGANISM bread wheat.

REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae
; Triticaceae; Triticum.
1 (bases 1 to 502)

AUTHORS Anderson, O.D., Chao, S., Han, P.S., Hsia, C.C., Johnson, R.R., Kang, Y.,
Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L., Tong, J.C., Verhey
S.D. and Walker-Simmons, M.K.
The structure and function of the expressed portion of the wheat
genomes - ABA-treated embryo library
Unpublished (2000)

JOURNAL Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105395818

Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Clontech Matchmaker 3' AD primer.
Location/Qualifiers

FEATURES

SOURCE

1..502
/organism="Triticum aestivum"
/cultivar="Brevor (soft, white, winter, common wheat)"
/db_xref="taxon:4565"
/clone="WHE0606_B07_C14"
/clone_lib="Wheat ABA-treated embryo cDNA library"
/tissue_type="Seed embryo"
/dev_stage="Mature dormant seeds"
/lab_host="E. coli DH12S"
/note="Vector: pGAD10; Site_1: EcoRI; Site_2: XhoI;
Embryos were cut from mature, dormant seeds and imbedded in
25 microm ABA (abscisic acid) in 5 mm Mes buffer, pH 5.7,
for 12 hr at 22 C. The tissue, total RNA, and poly(A)
RNA were prepared by Steven Verhey in M.K.
Walker-Simmons's lab (USDA-ARS, Washington State Univ.,
Pullman, Washington 99164-6420. A cDNA library was made
by Clontech using a combination of random and oligo dr
primers. Library was plated and archived by Russell
Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid
DNA preparations and DNA sequencing were performed in the
OD Anderson lab (all other authors)."
BASE COUNT 103 a 108 c 128 g 162 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 502;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgtctgttgaagc 18
|||||
Db 189 ATCTGCTGTTGAAGC 204

RESULT 7

LOCUS AO692054 517 bp DNA GSS 06-JUL-1999
DEFINITION HS_5408_A2_E07_T7A_RP11 Human Male BAC Library Homo sapiens
genomic clone Plate=984 Col=14 Row=L, DNA sequence.

ACCESSION AO692054
VERSION AO692054.1 GI:5382302
KEYWORDS GSS.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 517)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Mahairas, G.G., Wallace, J.C., Young, J., Zhao, S., Adams, M.D. and
Hood, L.

AUTHORS Sequencing-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RP11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 984 row: 1 column: 14
Seq primer: 97
Class: BAC ends
High quality sequence stop: 517.
Location/Qualifiers

FEATURES

SOURCE

1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=984 Col=14 Row=L"
/clone_lib="RP11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT 154 a 96 c 83 g 178 t 6 others
ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 517;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgtctgttgaagc 18
|||||
Db 411 ATCTGCTGTTGAAGC 426

RESULT 8

LOCUS AO889100 555 bp DNA GSS 10-NOV-1999
DEFINITION HS_2204_AL_C07_MR_C1T Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2204 Col=13 Row=E, DNA sequence.

ACCESSION AO889100
VERSION AO889100.1 GI:6345290
KEYWORDS GSS.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS        1 (bases 1 to 555)
                Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
                Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
                Hood,L.
TITLE          Sequence-tagged connectors: A sequence approach to mapping and
                scanning the human genome
JOURNAL        Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE        99380589
COMMENT        Contact: Mahairas GG, Wallace JC, Hood L
                High Throughput Sequencing Center
                University of Washington
                401 Queen Anne Avenue North, Seattle, WA 98109, USA
                Tel: (206) 616-3618
                Fax: (206) 616-3887
                Email: jwallace@u.washington.edu
                Clones may be purchased from Research Genetics (info@resgen.com).
                BAC end Web Server: http://www.hvsc.washington.edu
                Plate: 2204 row: E column: 13
                Seq primer: M13 Reverse
                Class: BAC ends
                High quality sequence stop: 555.
FEATURES
SOURCE
1..555
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2204 Col=13 Row=E"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelbac11; BAC clones in
E-Coli DH10B"
BASE COUNT      131 a 135 c 147 g 124 t 18 others
ORIGIN
Query Match      64.0%; Score 16; DB 13; Length 555;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4 tctgctgttgaaagc 19
|||||
Db 497 tctgctgttgaaagc 482

RESULT 9
AM064984 558 bp mRNA EST 12-OCT-1999
LOCUS ST38C02, mRNA sequence.
DEFINITION ST38C02, mRNA sequence.
ACCESSION AM064984
VERSION AM064984.1 GI:6020056
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
REFERENCE 1 (bases 1 to 558)
AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,F. and Sederoff,R.R.
TITLE The Pine Gene Discovery Project
JOURNAL Unpublished (1999)
COMMENT Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
, NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhetten@ncsu.edu
Seq primer: 5' lambda.TripLex2 Sequencing Primer.
FEATURES
SOURCE
1..558
/organism="Pinus taeda"
/db_xref="taxon:3352"

```

```

/clone="ST38C02"
/clone_lib="Pine TripLex shoot tip library"
/lab_host="E. coli BM25.8"
/note="Organ: shoot tips; Vector: lambda triplex; Site_1:
Site_1(A); Site_2: SfiI (B); Shoot tips (approx. 2 cm from
apex) were collected during the spring, frozen and used
for mRNA isolation. The SMART-PCR method (Clontech) was
used to prepare a library from 1 ug total RNA, using the
lambda triplex vector. Plasmid subclones in triplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."
BASE COUNT      146 a 117 c 131 g 160 t 4 others
ORIGIN
Query Match      64.0%; Score 16; DB 10; Length 558;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 gtttgaagcagcagc 25
|||||
Db 108 gtttgaagcagcagc 123

RESULT 10
BE232412/c 567 bp mRNA EST 10-JUL-2000
LOCUS BE232412/c
DEFINITION 137424 MARC IP1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE232412
VERSION BE232412.1 GI:9017130
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 567)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.M.
and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR primers
FORWARD: AGGAACAGCCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 72 row: E column: 24
Seq primer: ATTTAGTCACACTATAG.
FEATURES
SOURCE
1..567
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC IP1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT      114 a 198 c 171 g 83 t 1 others
ORIGIN
Query Match      64.0%; Score 16; DB 10; Length 567;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 8 ctgttgaagcagc 23
 |||||||
 Db 227 CTGTTTGAAGCGCAGC 212

RESULT 11
 AZ449344 632 bp DNA GSS 04-OCT-2000
 LOCUS
 DEFINITION clone UUGC1M0247J04 R. DNA sequence.
 ACCESSION AZ449344
 VERSION AZ449344.1 GI:10603038
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 632)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Irlam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.,
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112 USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0247 row: J column: 04
 Seq primer: CACGAGGAACACGTATGACC
 Class: plasmid ends
 High quality sequence stop: 632.
 Location/Qualifiers
 1..632
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0247J04"
 /clone_1lb="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1 resistant, F-"
 /note="vector: pMD42ny. Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adapted DNA was purified and agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g114732114[gb|AF129072.1]), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adapted mouse DNA was annealed to
 adapted vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 185 a 167 c 110 g 170 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 13; Length 632;
 Best Local Similarity 100.0%; Pred. No. 70;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggaatcgtgtttaa 16
 |||||||
 Db 552 GCATCTGCTGTTGAAC 567

RESULT 12
 TA69C05P/C 650 bp DNA GSS 13-DEC-2000
 LOCUS
 DEFINITION T. brucei sheared genomic DNA clone 69c05, forward sequence,
 genomic survey sequence.
 ACCESSION AL457556
 VERSION AL457556.1 GI:11858782
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei.
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.
 1 (bases 1 to 650)
 Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
 Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
 Melville, S.E., Rajadream, M.A. and Barrell, B.G.
 Direct Submission
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
 nh1@sanger.ac.uk
 Constructed at the Institute for Genomic Research (TIGR),
 Rockville, MD. Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 Gnat 10.1) was mechanically sheared
 to give a tight size distribution (4 kb). The v + 1 method used for the library construction is
 described in detail in Smith, H. and Venter, J.C. (Making small
 insert libraries for whole genome shotgun sequencing projects. In
 Genome Sequencing: A Practical Approach, eds. M. Vaubin and B.
 Barrell, Oxford University Press, 1999).
 Email: nelsayed@tigr.org
 Details of T. brucei sequencing at the Sanger Centre are available
 at <http://www.sanger.ac.uk/projects/T-brucei/>.
 Location/Qualifiers
 1..650
 /organism="Trypanosoma brucei"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="69c05"
 BASE COUNT 208 a 186 c 173 g 83 t
 ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 650;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 atctgctgttgaac 18
 |||||||
 Db 357 ATCTGCTGTTGAAC 342

RESULT 13
 FR0044714 177 bp DNA GSS 22-OCT-1999
 LOCUS
 DEFINITION Fugu rubripes GSS sequence, clone 192G14eff, genomic survey
 sequence.
 ACCESSION AL132206
 VERSION AL132206.1 GI:6114152
 KEYWORDS GSS; genome survey sequence.
 SOURCE Takifugu rubripes.
 ORGANISM Takifugu rubripes.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 177)
 AUTHORS Elgar, G., Clark, M.S., Smith, S., Meek, S., Warner, S., Edwards, Y.J.K.,
 Umranta, Y., Williams, G. and Brenner, S.
 TITLE Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
 JOURNAL Direct Submission
 Centre, Hinxton, Cambridge, CB10 1SB, UK Email:
 biohelp@hmp.mrc.ac.uk
 Vector: pBluescript II KS
 V-type: phagemid
 PRIMER: KS

COMMENT DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.
 FEATURES Location/Qualifiers
 source 1..177
 /organism="Hakifugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 192G14"
 /clone="192G14eb5"

BASE COUNT 33 a 41 c 50 g 44 t 9 others
 ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 177;
 Best Local Similarity 100.0%; Pred. NO. 2.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 atctgcgtttgaag 17
 |||||||
 Db 116 ATCTGCTTTGAAG 130

RESULT 14
 LOCUS BI024308 190 bp mRNA EST 14-JUN-2001
 DEFINITION CM3-MT0293-260101-692-f03 MT0293 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BI024308
 VERSION BI024308.1 GI:14430938
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 190)
 Dias Neto, F., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM3&cl2=CM3-MT0293-
 260101-692-f03&cl3=2001-01-26&cl4=1)
 Seg primer: puc 18 forward
 High quality sequence start: 80
 High quality sequence stop: 175.
 Location/Qualifiers
 1..190
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

FEATURES
 source

/clone_lib="MT0293"
 /dev_stage="Adult"
 /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (O.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."
 BASE COUNT 22 a 49 c 64 g 54 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 190;
 Best Local Similarity 100.0%; Pred. NO. 2.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 ctgcgtttgaagc 19
 |||||||
 Db 44 CTGCTTTGAAGCG 58

RESULT 15
 LOCUS AI021028/c 278 bp mRNA EST 16-JUN-1998
 DEFINITION u099b01.r1 Soares_mammary_gland_NbMNG Mus musculus cDNA clone
 IMAGE:1365577 5' similar to TR:035626 035626 RAS-RELATED PROTEIN.
 ;, mRNA sequence.
 ACCESSION AI021028
 VERSION AI021028.1 GI:3235364
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 278)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Stepoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, R., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:898797

FEATURES POSSIBLE REVERSED CLONE: Similarity on wrong strand
 Seg primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 106.
 Location/Qualifiers
 1..278
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1365577"
 /clone_lib="Soares_mammary_gland_NbMNG"
 /sex="male"
 /tissue-type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: mammary gland; Vector: p7T73D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)
 primer [5',
 TGTACCAATCTGAGTGGAGGCGCGCAATGCTTTTCTTTTCTTTTCTTTTCTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 316;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctcgtgttgaag 17
|||||

Db 149 ATCTGCTGTTCGAG 135

RESULT 18
BF355075 358 bp mRNA EST 22-NOV-2000
LOCUS RC2-HT0827-240700-014-h01 HT0827 Homo sapiens cDNA, mRNA sequence.
DEFINITION
ACCESSION BF355075
VERSION BF355075.1 GI:11314149
KEYWORDS
SOURCE human.
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 358)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M. R.,
Negal, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC2&f2=RC2-HT0827-
240700-014-h01&t3=2000-07-24&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 13
High quality sequence stop: 358.

FEATURES
source
1. 358
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0827"
/dev_stage="Adult"
/note="Organ: head-neck; Vector: puc18; site_1: SmaI;
site_2: SmaI; A mini-library was made by cloning products
derived from ORESPEC PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 88 a 111 c 81 g 78 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 358;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctcgtgttgaag 18
|||||

Db 275 TCTGCTGTTCGAGC 261

RESULT 19

A0521765 372 bp DNA GSS 05-MAY-1999
LOCUS HS_5239_A2_H09_SBE6 RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION
ACCESSION A0521765
VERSION A0521765.1 GI:4752890
KEYWORDS
SOURCE human.
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 372)
Mahairas, G. G., Wallace, J. C., Smith, K., Swartzel, S., Holzman, P.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D., and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering-bac.htm)
or from Resear h Genetics (Info@resgen.com). BAC end Web Server:
http://www.husc.washington.edu
Plate: 815 row: O column: 18
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 372.

FEATURES
source
1. 372
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=815 Col=18 Row=O"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="Male"
/note="Vector: pBACe3.6; site_1: EcoRI; site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

BASE COUNT 97 a 86 c 72 g 115 t 2 others
ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gactcgtgttgaag 16
|||||

Db 315 GATGCTGTTCGAG 329

RESULT 20
A0143500/c 406 bp DNA GSS 24-SEP-1998
LOCUS HS_3066_B2_H10_MR_CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3066 Col=20 Row=P, DNA sequence.
ACCESSION A0143500
VERSION A0143500.1 GI:3534153
KEYWORDS
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 406)
MAHAIKAS, G.G., WALLACE, J.C., SMITH, K., SWARTZELL, S., HOLZMAN, T.,
KELLER, A., SHAKER, R., FURLONG, J., YOUNG, J., ZHAO, S., ADAMS, M.D. and
HOOD, L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: MahaiRAS GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence tagged connector
Plate: 3066 row: P column: 20
Class: BAC ends
High quality sequence stop: 406.
Location/Qualifiers
1..406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3066 Col=20 Row=P"
/clone_lib="CIR Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAcl1; BAC Clones in
E-Coli DH10B"
BASE COUNT 138 a 78 c 54 g 135 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 406;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgttgaag 16
|||||
Db 98 GATCTGCTGTGAAG 84

RESULT 21
AQ494880 424 bp DNA GSS 28-APR-1999
LOCUS
DEFINITION HS_5195_A1.B06.SPEE RPCR-11 Human Male BAC library Homo sapiens
genomic clone Plate=771 Col=11 Row=C, DNA sequence.
ACCESSION AQ494880
VERSION AQ494880.1 GI:4695047
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 424)
MAHAIKAS, G.G., WALLACE, J.C., SMITH, K., SWARTZELL, S., HOLZMAN, T.,
KELLER, A., SHAKER, R., FURLONG, J., YOUNG, J., ZHAO, S., ADAMS, M.D. and
HOOD, L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: MahaiRAS GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCR-11. For BAC
library availability, please contact Pieter de Jong

(pieter@ejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (<http://www.htsc.washington.edu>)
<http://www.htsc.washington.edu>
Plate: 771 row: C column: 11
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 424.
Location/Qualifiers
1..424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=771 Col=11 Row=C"
/clone_lib="RPCR-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site: 1: EcoRI; Site: 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 137 a 70 c 82 g 134 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 424;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaagc 18
|||||
Db 91 TCTGCTGTGAAGC 105

RESULT 22
BH077581 427 bp DNA GSS 18-JUL-2001
LOCUS
DEFINITION RPCR-24-291N2.TJ RPCR-24 Mus musculus genomic clone RPCR-24-291N2,
DNA sequence.
ACCESSION BH077581
VERSION BH077581.1 GI:14897178
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 427)
ZHAO, S., NIERMAN, W., MALEK, J., SHATSMAN, S., AKINRE, B., LEVINS, M.,
TSEGAYE, G., GEET, K., KROL, M., SHVARTSBEYN, A., GEORGEORGIS, E.,
RUSSELL, D., DE JONG, P. and FRASER, C.M.
Mouse BAC End Sequences from Library RPCR-24
Unpublished (1999)
Other GSSs: RPCR-24-291N2.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaod@igf.org
Clones are derived from the mouse BAC library RPCR-24. For BAC
library availability, please contact Pieter de Jong
BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 291 row: N column: 2
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..427
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCR-24-291N2"

```

/clone_lib="RPC1-24"
/sex="Male"
/Cell_type="Spleen/Brain"
/Note="Vector: pFARBAC1; Site_1: BamHI; Site_2: BamHI;
RPC1-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pFARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57Bl/6J
DNA."
BASE COUNT      102 a      106 c      93 g      126 t
ORIGIN

Query Match      60.0%; Score 15; DB 13; Length 427;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgcgtttgaagc 18
|||||
Db 326 TCTGCTGTTGAAGC 340

RESULT 23
LOCUS      A0114851      435 bp      DNA      GSS      29-AUG-1998
DEFINITION CIT-HSP-2376L11.TF CIT-HSP Homo sapiens genomic clone 2376L11, DNA
sequence.
ACCESSION  A0114851
VERSION    A0114851.1 GI:3490972
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 435)
AUTHORS    Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Henry,K., Gaenger,D., Sun,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
            Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
            Unpublished (1998)
            Other_GSS: CIT-HSP-2376L11.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (Info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES
    source
        1..435
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="2376L11"
            /clone_lib="CIT-HSP"
            /sex="Male"
            /cell_type="Sperm"
            /note="Vector: pFARBAC1; Site_1: HindIII; Site_2:
            HindIII"
BASE COUNT      87 a      115 c      118 g      115 t
ORIGIN

Query Match      60.0%; Score 15; DB 13; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatcgcgtttgaa 16
|||||

```

```

Db 375 CATCTGCTGTTTGA 389

RESULT 24
LOCUS      AM432846      437 bp      mRNA      EST      11-FEB-2000
DEFINITION S181905.Y1 Gm-c1016 Glycine max cDNA clone GENEOME SYSTEMS CLONE ID:
Gm-c1016-6609 5' similar to TR:Q40171 Q40171 TDR6 MRNA ;, mRNA
sequence.
ACCESSION  AM432846
VERSION    AM432846
KEYWORDS   EST.
SOURCE     soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE  1 (bases 1 to 437)
AUTHORS    Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
            A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            R., Ritter,E., Kohn,S., Shin,T., Jackson,T., Cardenas,M., McCann
            R., Waterston,R. and Wilson,R.
            Public Soybean EST Project
            Unpublished (1999)
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 for further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Glibco
            High quality sequence stop: 350.
FEATURES
    source
        1..437
            /organism="Glycine max"
            /db_xref="taxon:3847"
            /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-6609"
            /clone_lib="Gm-c1016"
            /tissue_type="Immature flowers of field grown plants"
            /lab_host="X110-Gold"
            /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
            XhoI; This cDNA library was constructed from mRNA isolated
            from immature flowers of field grown plants. The cDNA
            library was prepared using the Stratagene pBluescript II
            XR library construction kit. Complementary DNA was
            synthesized from mRNA using a primer consisting of a poly
            (dT) sequence with a XhoI restriction site. EcoRI adapters
            were ligated to the blunt-ended cDNA fragments followed by
            XhoI digestion. The cDNA fragments were directionally
            cloned into the EcoRI-XhoI restriction site of the
            pBluescript vector. The ligated cDNA fragments were
            transformed into XL10-Gold host cells. This library was
            constructed by Dr. Randy Shoemaker and Dr. John
            Erpelting."
BASE COUNT      133 a      101 c      83 g      120 t
ORIGIN

Query Match      60.0%; Score 15; DB 10; Length 437;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatcgcgtttgaa 16
|||||

```

Db 123 GATCTGCTTTTGA 137

RESULT 25
LOCUS BF555582/c
DEFINITION BF555582 456 bp mRNA EST 12-DEC-2000
UT-R-EO-cy-d-07-0-01.1 UT-R-EO Rattus norvegicus cDNA clone
ACCESSION BF555582
VERSION BF555582.1 GI:11665312
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 456)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LIML (limlimage.liml.gov). IMAGE ID= 1770501
Seq primer: M13 forward

FEATURES
Source
Location/Qualifiers
1..456
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-EO-cy-d-07-0-01"
/clone_1ib="UI-R-EO"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."

BASE COUNT 112 a 115 c 139 g 88 t 2 others
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggactcgtgttga 15
|||||
Db 88 GGATCTGCTTTTGA 74

RESULT 26
LOCUS A1477772/c
DEFINITION A1477772 466 bp mRNA EST 07-JUN-2001
ID54h1.y1 zebrafish Washu MPMG EST Danio rerio cDNA clone
IMAGE:3715749 5', mRNA sequence.
ACCESSION A1477772
VERSION A1477772.1 GI:4335383
KEYWORDS EST.
SOURCE zebrafish.

ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbortinae; Danio.
REFERENCE 1 (bases 1 to 466)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
S., Hillier,L., Knab,T., Martin,D., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Pearson,B.,
Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
TITLE Washu zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResourceZentrumPirmasens, Berlin, Germany (web address:
www.rzp.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 465
POLYA-No.

FEATURES
Source
Location/Qualifiers
1..466
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="IMAGE:3715749"
/clone_1ib="Zebrafish Washu MPMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="X11-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'gGACTGATTCAGATCGGAGCGCCGCCCTTTTCTTTTCTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach Lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

BASE COUNT 97 a 135 c 120 g 113 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 466;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 ctgttgaagcag 22
|||||
Db 330 CTGTTTGAAGCGCAG 316

RESULT 27
A2965014

LOCUS A2965014 479 bp DNA GSS 27-APR-2001
 DEFINITION 2M0234K24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 clone UUGC2M0234K24 R, DNA sequence.
 ACCESSION A2965014
 VERSION A2965014.1 GI:13836241
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 479)
 REFERENCE Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relliy,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 TITLE Unpublished (2000)
 JOURNAL
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 std Error: 0.00
 Plate: 0234 row: K column: 24
 Seq primer: CACACACGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 479.
 Location/Qualifiers
 1. 479
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0234K24"
 /clone_1lb="Mouse 10kb plasmid UUGC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD29uv. Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (g11473211419b/AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 163 a 91 c 122 g 103 t
 ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 479;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatctgctgtttgaa 16
 ||||||||||||||||
 Db 197 GATCTGCTGTTTGA 211

RESULT 28

AO192174
 LOCUS AO192174 495 bp DNA GSS 04-NOV-1998
 DEFINITION HS-3227_B1_G01_T7 CTF Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=3227 Col=1 Row=N, DNA sequence.
 ACCESSION AO192174
 VERSION AO192174.1 GI:3590796
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 495)
 REFERENCE Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
 Hood, L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589
 JOURNAL
 MEDLINE
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3227 Row: N column: 1
 Class: BAC ends
 High quality sequence stop: 495.
 Location/Qualifiers
 1. 495
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3227 Col=1 Row=N"
 /clone_1lb="CTF Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="organ: sperm; Vector: pBelBAC11; BAC clones in
 E-Coli DH10B"

BASE COUNT 135 a 114 c 79 g 167 t
 ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 495;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatctgctgtttgaa 16
 ||||||||||||||||
 Db 316 GATCTGCTGTTTGA 330

RESULT 29

BEL14310/c
 LOCUS BEL14310/c 507 bp mRNA EST 13-JUN-2000
 DEFINITION UI-R-BJ1-gwk-g-02-0-UI-s1 UI-R-BJ1 Rattus norvegicus cDNA clone
 UI-R-BJ1-gwk-g-02-0-UI-3', mRNA sequence.
 ACCESSION BEL14310
 VERSION BEL14310.1 GI:8506415
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 507)
 REFERENCE Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping

University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized heart library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

Location/Qualifiers
1..507
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ1-awk-9-02-0-01"
/clone_lib="UI-R-BJ1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1 library is a subtracted library derived from the following tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, ventricle at 13 dpc, and adult heart. For a detailed description of the library from which this clone was derived, please visit our web site at ratseq.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-BJ1
TAG_TISSUE=heart
TAG_SRO=ACAC
BASE COUNT 125 a 119 c 105 g 158 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 507;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgtttgaagc 18
|||||
Db 357 TCTGCTGTTGAAGC 343

RESULT 30
AW915326 509 bp mRNA EST 25-MAY-2000
LOCUS EST346630 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
DEFINITION R61C01 5' end, mRNA sequence.
ACCESSION AW915326
VERSION AW915326.1 GI:8081017
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 509)
AUTHORS Lee, N.H., Glodde, A., Chandra, I., Mason, T.M., Quackenbush, J., Kerlavage, A.R. and Adams, M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529

Fax: (301)-838-0208
Email: nhlee@igf.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information
Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers
1..509
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="R61C01"
/clone_lib="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pT73D-pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT 116 a 167 c 118 g 107 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 509;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtgtttga 15
|||||
Db 443 GCATCTGCTGTTGA 429

RESULT 31
AA818607 528 bp mRNA EST 03-JUL-1999
LOCUS AA818607/c
DEFINITION UI-R-A0-bc-h-05-0-01.s1 UI-R-A0 Rattus norvegicus cDNA clone
UI-R-A0-bc-h-05-0-01 3' similar to gb|U96683|MMU96683 Mus musculus immunoglobulin-like receptor PIRNA2 (6M23) mRNA, partial cds, mRNA sequence.

ACCESSION AA818607
VERSION AA818607.1 GI:4228398
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 528)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT On Feb 17, 1998 this sequence version replaced gi:2889346.
CONTACT: Soares, MB
PROGRAM FOR RAT GENE DISCOVERY AND MAPPING
UNIVERSITY OF IOWA
451 ECKSTEIN MEDICAL RESEARCH BUILDING IOWA CITY, IA 52242, USA
TEL: 319 335 8250
FAX: 319 335 9565
EMAIL: msares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult kidney library. cDNA library preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE ID-1768272
Seq primer: M13 Forward
POLYA=No.

FEATURES

source

Location/Qualifiers
1..528
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-bc-h-05-0-01"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified

polylinker. Site.1: Not I; Site.2: Eco RI. This library consists of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 138 a 131 c 101 g 158 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 tctgtctgttgaagc 18
|||||
Db 353 TCTGCTGTTGAAGC 339

RESULT 32
TA375E1P/c 532 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 375e11, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL495591 GI:11873033
VERSION GSS.
KEYWORDS Trypanosoma brucei.
SOURCE Trypanosoma brucei
ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 532)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.
DIRECT SUBMISSION
COMMENT Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhls@sanger.ac.uk
CONSTRUCTED at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (FREU927/4 G9at 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + 1 method used for the library construction is described in detail in Smith, H. and Venler, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at <http://www.sanger.ac.uk/projects/T-brucei/>.
Location/Qualifiers

FEATURES
source 1..532
/organism="Trypanosoma brucei"
/strain="FREU927"
/db_xref="taxon:5691"
/clone="375e11"

BASE COUNT 162 a 156 c 92 g 122 t
ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 532;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 ctgtctgttgaagc 19
|||||
Db 310 CTGCTGTTGAAGC 296

RESULT 33

AA790463/c 542 bp mRNA EST 06-FEB-1998
LOCUS AA790463
DEFINITION v040412.t1 Soares_mammary_gland_NbMG Mus musculus cDNA clone
IMAGE:1230982.5 similar to SW:RAP_HUMAN P17964 RAS-RELATED
PROTEIN RAP-2B. // mRNA sequence.

ACCESSION AA790463
VERSION AA790463.1 GI:2850583
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 542)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellendberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
TITLE The WashU-HM Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HM Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:656574
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 492.
Location/Qualifiers

FEATURES
source 1..542

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1230982"
/clone_id="Soares_mammary_gland_NbMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: p7T3D-Pac (Pharmacia RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5'-TGTTACCAATCTGAAGTGGAGCGCCGCAATGTTTTTTTTTTTTTTTTTTT T 3'-); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 127 a 178 c 125 g 112 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 542;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggaatctgttga 15
|||||
Db 486 GGATCTGCTTTGA 472

RESULT 34
TA162C10P/c 559 bp DNA GSS 13-DEC-2000
LOCUS TA162C10P
DEFINITION T. brucei sheared genomic DNA clone 162c10, forward sequence,
genomic survey sequence.
ACCESSION AL472440

SOURCE
ORGANISM
Takifugu rubripes.
Takifugu rubripes

REFERENCE
AUTHORS
Elgar, G., Clark, M.S., Smith, S., Meek, S., Warner, S., Edwards, Y.J.K., Umranta, Y., Williams, G. and Brenner, S.
Direct Submission
Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email: biohelp@hmp.mrc.ac.uk
Vector: pBluescript II KS
V-type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic sequence.

FEATURES
Location/Qualifiers
1..600
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 192G14"
/clone="192G14bC7"

BASE COUNT
123 a 152 c 153 g 170 t 2 others

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 600;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgcctgttgaag 17
|||||

Db 145 ATCTGCTTTGAAG 159

RESULT 38
FR0044632/c
LOCUS
Definition
Fugu rubripes GSS sequence, clone 192G14bC7, genomic survey sequence.
Accession
AL132124
Version
AL132124.1 GI:6114070
Keywords
GSS: genome survey sequence.
Source
Takifugu rubripes.
Organism
Takifugu rubripes

REFERENCE
AUTHORS
Elgar, G., Clark, M.S., Smith, S., Meek, S., Warner, S., Edwards, Y.J.K., Umranta, Y., Williams, G. and Brenner, S.
Direct Submission
Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email: biohelp@hmp.mrc.ac.uk
Vector: pBluescript II KS
V-type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic sequence.

FEATURES
Location/Qualifiers
1..611
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 192G14"
/clone="192G14bC7"

BASE COUNT
158 a 154 c 150 g 122 t 27 others

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 611;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgcctgttgaag 17
|||||

Db 450 ATCTGCTTTGAAG 436

RESULT 39
BG306774/c
LOCUS
Definition
Fugu rubripes GSS sequence, clone 192G14bC7, genomic survey sequence.
Accession
BG306774
Version
BG306774.1 GI:13104301
Keywords
GSS: genome survey sequence.
Source
Takifugu rubripes.
Organism
Takifugu rubripes

REFERENCE
AUTHORS
Clark, M., Johnson, S.L., Lehnach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stephe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Wasnu zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@wustl.edu
Library constructed by: Susan E. Brockerhoff DNA Sequencing by:
Washington University Genome Sequencing Center Clone distribution:
ResourceCenter@umrprimatdatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence scop: 420.

FEATURES
Location/Qualifiers
1..621
/organism="Danio rerio"
/strain="wild-type"
/db_xref="taxon:7955"
/clone="414535"
/clone_lib="Zebrafish adult retina cDNA"
/sex="mixed"
/dev stage="1-2 years"
/lab host="E. Coli XL1-Blue MRF"
/note="Vector: Lambda ZAP II (pBluescript SK-); Site 1:
ECORI; Site 2: SalI; This Zebrafish library was
constructed by Dr. Susan E. Brockerhoff (email:
sbrocker@u.washington.edu) RZPD library number: 760"

BASE COUNT
157 a 141 c 164 g 154 t 5 others

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 621;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cgtttgaagcgag 22
|||||

Db 180 CTGTTGAAGCGAG 166

```

RESULT 40
LOCUS BG017095/c 627 bp mRNA EST 18-APR-2001
DEFINITION df68g93.x1 Xenopus laevis unfertilized egg cDNA library Xenopus
laevis cDNA clone IMAGE:3744580 3', mRNA sequence.
ACCESSION BG017095
VERSION BG017095.1 GI:12470841
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 627)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Page,D.,
Martin,J., Wyllie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Other ESTs: df68g03.y1
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: http://www.resgen.com/ Please reference the id listed
below when ordering this clone: Source lab clone id - xlneg005n05
Seq primer: -40up from Gibco
High quality sequence stop: 468.
Location/Qualifiers
1..627
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3744580"
/clone_lib="Xenopus laevis unfertilized egg cDNA library"
/tissue_type="unfertilized egg"
/lab_host="Top-10 F'"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from 2ug of poly A+ RNA.
EcoRI-XhoI cut cDNA was then ligated into XhoI-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the 3'
end. SS-library phagemids were prepared by mass excision
from the original library and normalized by hybridization
to biotinylated driver (prepared from the same library by
PCR) to Cot-omega of 11. After removal of hybrids and
excess driver by streptavidin sepharose chromatography,
the ss-phagemids were made double stranded and
electroporated into Top-10 F'. Original library
construction by Bruce Blumberg (Blumberg et al., 1991
Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
2923-2935). Note: This is a Xenopus Gene Collection (XGC)
library."
BASE COUNT 146 a 131 c 143 g 206 t 1 others
ORIGIN
Query Match 60.0%; Score 15; DB 11; Length 627;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 11 tttaagcgcagcga 25
DB 556 TTTGAAGCGCAGCAG 542

```

```

RESULT 41
LOCUS AO754510 632 bp DNA GSS 27-JUL-1999
DEFINITION HS_5434_AL-G05_SP6E_RPCI-11 Human Male BAC library Homo sapiens
genomic clone SP6E-1010 Col-9 Row-M, DNA sequence.
ACCESSION AO754510
VERSION AO754510.1 GI:5618726
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 632)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
JOURNAL
MEDLINE
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@edlong.med.buffalo.edu). Clones may be purchased from
BACAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)
or from Research Genetics (Inforesgen.com). BAC end Web Server:
http://www.hsc.washington.edu
Plate: 1010 row: M column: 9
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 632.
Location/Qualifiers
1..632
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1010 Col=9 Row=M"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT 156 a 143 c 126 g 190 t 17 others
ORIGIN
Query Match 60.0%; Score 15; DB 13; Length 632;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 gctgttgaaagcga 21
DB 308 GCGTTTGAAGCGCA 322

```

```

RESULT 42
LOCUS BG546580 640 bp mRNA EST 04-APR-2001
DEFINITION 602574435E1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4702473 5',
mRNA sequence.
ACCESSION BG546580
VERSION BG546580.1 GI:13545245
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE 1 Mammalia: Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 (bases 1 to 640)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
 Plate: L10M1539 row: f column: 10
 High quality sequence stop: 640.

FEATURES
 Source Location/Qualifiers
 1..640
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4702473"
 /clone_lib="NIH MGC 77"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: Lung; Vector: pDR-LJB (Clontech); Site_1: S11 (ggccgcctcgcc); Site_2: S11 (ggccatcggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCCGCGCCGACATG-dr(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."

BASE COUNT 170 a 131 c 155 g 184 t
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 640;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atcgcgtcttgaag 17
 Db 600 ATCTGCTCTTGAAG 614

RESULT 43
 LOCUS BG969048 644 bp mRNA EST 12-JUN-2001
 DEFINITION 602835003F1 NCI_CGAP_Co24 Mus musculus CDNA clone IMAGE:4989312 5',
 mRNA sequence.
 ACCESSION BG969048
 VERSION BG969048.1 GI:14356685
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa;
 Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
 Plate: L10M1003 row: f column: 01
 High quality sequence stop: 642.

FEATURES
 Source Location/Qualifiers
 1..644
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4989312"
 /clone_lib="NCI_CGAP_Co24"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 153 a 196 c 164 g 131 t
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 644;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtcttga 15
 Db 470 GGAATCTCTTTGA 456

RESULT 44
 LOCUS BE597936 645 bp mRNA EST 18-AUG-2000
 DEFINITION P11_67_C08_g1A002 Pathogen induced 1 (P11) Sorghum bicolor CDNA,
 mRNA sequence.
 ACCESSION BE597936
 VERSION BE597936.1 GI:9853009
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 645)
 Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
 ,L.H.
 An EST database from Sorghum: pathogen-induced plants
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: PolyTmix
 High quality sequence start: 3
 High quality sequence stop: 635
 POLYA=No.

FEATURES
 Source Location/Qualifiers
 1..645
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen induced 1 (P11)"
 /note="Organ: Anthranose-infected leaves from
 two-week-old sorghum plants 48 hr after inoculation;
 Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
 Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
 cultivar) were infected with pathogen (isolate FRM421 of
 Colletotrichum graminicola, which is a sorghum isolate).
 RNA was prepared from infected leaves harvested from 45
 seedlings 48 hours after inoculation. Note: young
 seedlings (2 weeks old) exhibit juvenile resistant
 reaction, which is an incompatible interaction. As they
 grow older (4 weeks or older), plants resume susceptibility

to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 179 a 143 c 170 g 153 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 645;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 ctgcctttgaagc 19
|||||
Db 627 CTGCTGTTGAAGC 641

RESULT 45
BG085090 649 bp mRNA EST 26-JAN-2001
LOCUS H3108E05-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
DEFINITION H3108E05 5', mRNA sequence.
ACCESSION BG085090
VERSION BG085090.1 GI:12567654
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 649)
Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
,T.S., Carter,M.G. and Ko,M.S.H.
Verification and initial annotation of NIA mouse 15K cDNA clone set
Unpublished (2001)
Other ESTs: H3108E05-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit <http://igsun.grc.nia.nih.gov/cDNA/15k.html> for details.
Plate: H3108 row: E column: 05
Seq primer: -21M13 Reverse
High quality sequence stop: 649
POLYA-No.

FEATURES
Source
Location/Qualifiers
1..649
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="H3108E05"
/clone_id="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT) Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000,

Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 157 a 182 c 177 g 133 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 649;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggatcctgctgttga 15
|||||
Db 315 GGATCTGCTGTTGA 301

Search completed: December 26, 2001, 14:10:10
Job time: 7057 sec